

**Univerzita Karlova v Praze**

**Filozofická fakulta**

**Ústav lingvistiky a ugrofinistiky**

## **DIPLOMOVÁ PRÁCE**

**Milan Souček**

**Kontrastivní studie substantiv vyjadřujících čas ve finštině a  
češtině**

**Contrastive study of time nouns in Finnish and Czech**

**Tšekin ja suomen kielen ajan substantiivien kontrastiivinen analyysi**

**Vedoucí práce: PhDr. Hilikka Lindroos, CSc.**

## Poděkování

Chtěl bych tímto poděkovat všem členům oddělení ugrofinistiky Ústavu lingvistiky a ugrofinistiky FF UK v Praze, vedeného paní PhDr. Hilkkou Lindroos, CSc., za profesionální přípravu a přátelské prostředí, které mně i mým spolužákům poskytovali po celou dobu našeho studia. Mgr. Anně Čermákové chci poděkovat za odbornou pomoc a rady při psaní této práce. Svým rodičům, Eevě Peltonen a všem svým blízkým děkuji za vytrvalou podporu a pomoc, bez níž bych se při studiu na FF UK jen stěží obešel.

Věnováno B z B

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a že jsem uvedl všechny použité prameny a literaturu.

A handwritten signature in black ink, appearing to read 'Milan Souček'.

Milan Souček

## Obsah:

Úvod .....	8
1 Cíl práce.....	10
2 Teoretický rámec .....	13
2.1 Jazykový korpus.....	13
2.2 Zkoumání jazyka jako produktu komunikace .....	14
2.3 Kolokace.....	16
2.4 Metody <i>corpus-based</i> a <i>corpus-driven</i> .....	17
3 Metodologie a data .....	20
3.1 Substantiva vyjadřující čas.....	20
3.2 Kielipankki.....	21
3.3 Práce s WWW Lemmie 2.0.....	24
3.4 Český národní korpus.....	25
4 Korpusové analýzy .....	27
4.1 Statistiky.....	27
4.1.1 Statistické údaje.....	27
4.1.2 Pádová distribuce substantiv času ve finštině.....	30
4.1.3 Míra asociace.....	40
4.2 Funkční skupiny substantiv vyjadřujících čas ve finštině.....	43
4.2.1 Všeobecná substantiva.....	43
4.2.2 Skupina jednotek času.....	44
4.2.3 Postpoziční spojení.....	45
4.2.4 Spojení s adjektivem typu <i>mittainen</i> .....	47

4.2.5	Deiktické výrazy.....	47
4.2.6	Idiomatická spojení.....	49
5	Zpracování materiálů pro slovníková hesla.....	50
5.1	Zpracování materiálů.....	50
5.2	Finsko-český slovník.....	52
5.2.1	Substantivum <i>aika</i> .....	54
5.2.2	Substantivum <i>vuosi</i> .....	57
5.2.3	Substantiva <i>kuu</i> a <i>kuukausi</i> .....	58
5.2.4	Substantivum <i>viikko</i> .....	59
5.2.5	Substantivum <i>päivä</i> .....	60
5.2.6	Substantivum <i>tunti</i> .....	61
5.2.7	Substantiva <i>minuutti</i> a <i>sekunti</i> .....	62
5.3	Česko-finský slovník.....	62
5.3.1	Substantivum <i>čas</i> .....	64
5.3.2	Substantivum <i>rok</i> .....	65
5.3.3	Substantivum <i>měsíc</i> .....	66
5.3.4	Substantivum <i>týden</i> .....	66
5.3.5	Substantivum <i>den</i> .....	67
5.3.6	Substantivum <i>hodina</i> .....	67
5.3.7	Substantiva <i>minuta</i> , <i>sekunda</i> a <i>vteřina</i> .....	68
	Závěr.....	69
	Přílohy.....	73
	Seznam zkratk.....	99

Yhteenveto.....100

Literatura.....103

## Úvod

Lingvistika počátku 21. století má nové a rychle se rozvíjející možnosti a výzkumné nástroje, které jí dávají dříve nedosažitelné a téměř nepředstavitelné možnosti. Při zkoumání a popisu jazyka a následné přípravě jazykových příruček, učebnic a slovníků může využívat moderní technologie a jejich prostřednictvím zkoumat jazyk nejen jako systém dorozumívání, ale také jako výsledek komunikace, a na základě těchto zkoumání realizovaných jazykových projevů vytvářet teorie podložené konkrétními a přirozenými příklady. Již téměř samozřejmostí se stává, že moderní slovníky jsou připravovány na základě zkoumání obsáhlých jazykových materiálů nashromážděných v jazykových korpusech a zpracovávaných korpusovými programy, které umožňují zkoumání pestré škály jazykových jevů. Také při přípravě jazykových učebnic se jazykové korpusy využívají stále častěji a jazykové učebnice tak podávají studentům adekvátní a relevantní informace o jazyce, stále více nezávislé na subjektivním výběru jejich tvůrců.

Velké moderní slovníky a jazykové učebnice založené na korpusových zkoumáních vznikaly nejprve především v anglofonním prostředí (např. Collins Cobuild), v posledním desetiletí se ale korpusová lingvistika rozvíjí velkou rychlostí i v dalších zemích. Například Český národní korpus založený v roce 1994 může svým rozsahem a kvalitou směle konkurovat mnoha národním korpusům evropským i americkým. Publikace vytvořené na materiálech ČNK a vydané v posledních několika letech<sup>1</sup> ukazují, že i malé jazyky jako je čeština mají v současné době potenciál a možnosti být zkoumány a zpracovávány na základě moderních přístupů k jazyku a že

---

<sup>1</sup> Např. Frekvenční slovník češtiny. NLN. Praha 2004, Frekvenční slovník mluvené češtiny. Karolinum. Praha 2007, a další.



tato jazyková zkoumání mohou být úspěšně zhodnocena i při přípravě praktických jazykových publikací, slovníků a učebnic.

Problematika takzvaných malých jazyků a jejich prosazení v konkurenci jazyků mezinárodní komunikace (angličtiny, francouzštiny, španělštiny, v Evropě němčiny a nově se do popředí deroucí ruštiny) se zdá v prostředí Evropské unie být alespoň zčásti vyřešena rovnoprávným postavením jazyků jednotlivých členských států EU a jejich automatickým statutem úředních jazyků Evropské unie. Odborná příprava překladatelů a tlumočnicků především mezi tzv. malými jazyky vyžaduje rozvoj studia těchto jazyků v jednotlivých zemích EU a také s tím související přípravu materiálů k výuce těchto jazyků. Proto například na mnoha ústavech a katedrách evropských jazyků FF UK v Praze vznikají nové slovníky, pro jejichž tvorbu se v současné době využívá už i materiálů získaných z korpusových výstupů, například v rámci projektu Intercorp<sup>2</sup>. Vedle využití paralelních korpusů je další možností využití kontrastivních analýz jazyků na základě jejich národních korpusů. V této práci se chceme pokusit o takovýto druh analýzy češtiny a finštiny a získané výsledky a materiály chceme využít jako materiál pro tvorbu nového překladového dvojjazyčného slovníku těchto dvou jazyků.

---

<sup>2</sup> Více o projektu Intercorp v kapitole 1.

# 1 Cíl práce

Cílem této práce je metodami korpusové lingvistiky prozkoumat lexikální obraz vybrané skupiny lexémů v rámcich kontextuálních a funkčně-sémantických (Firth 1957, Sinclair 1991) a z nich vycházejících korpusově lingvistických přístupů k jazyku a prověřit možnosti využití korpusů českého a finského jazyka pro kontrastivní studium finštiny a češtiny, které by bylo možné zhodnotit například při tvorbě nového překladového dvojjazyčného finsko-českého slovníku. Na skupině substantiv vyjadřujících čas chceme s pomocí dat získaných z Českého národního korpusu (ČNK), dostupného online na WWW stránkách Ústavu Českého národního korpusu Filozofické Fakulty Univerzity Karlovy v Praze (ÚČNK FF UK)<sup>3</sup>, a finského jazykového korpusu Kielipankki, dostupného online na WWW stránkách Tieteen tietotekniikan keskus CSC<sup>4</sup>, prozkoumat využitelnost obou korpusů pro kontrastivní lexikologické a lexikografické analýzy obou jazyků a zjistit možnosti kombinovatelnosti výsledků získaných analýzou materiálů obou korpusů. Pro svou analýzu jsme si vybrali skupinu substantiv vyjadřujících čas (čas/aika, rok/vuosi, měsíc/kuu(kausi), den/päivä, hodina/tunti, minuta/minuutti, sekunda/vteřina/sekunti).

Vybraná slova chceme podrobit mnohostranné korpusové analýze – zaměříme se na jejich frekvenci v rámci zkoumaného textového materiálu, na jejich funkčně-sémantický obraz a svou pozornost upřeme především na jejich zapojení do jazykových struktur a kontextů – na jejich kolokabilitu, spojitelnost s jinými lexikálními jednotkami. Jedině ve svých kontextech užití, a to jak v lingvistických, tak v sociálně-kulturních dimenzích, nabývají lexémy svých aktuálních významů a jedině sledování jejich funkce v kontextu nám může podat celistvý obraz o proměnách a fungování

---

<sup>3</sup> <http://ucnk.ff.cuni.cz/>

<sup>4</sup> <http://www.csc.fi/tutkimus/alat/kielitiede>, oficiální název instituce v angličtině: Finnish IT Center for Science.

těchto významů. Vytržením lexémů z kontextu jejich užití se omezuje realizovatelnost jejich významů a funkcí, a proto je nezbytné je sledovat v rámci jejich sémantických vztahů v kontextu. Takovéto sledování umožňuje korpusová lingvistika, která je schopna na základě velkých textových materiálů vyhledat a zobrazit lexémy v jejich kontextovém zapojení.

Lexikografická práce dlouhá desetiletí spočívala ve shromažďování excerpčních lístků v lexikografických centrech a sběru příkladů užití slov z krásné literatury, nebo dokonce vymyšlení příkladů vlastních. Využití moderních technologií v lingvistice, vznik jazykových korpusů a možnost zkoumání nesrovnatelně obsáhlejších textových materiálů v neporovnatelně kratší době nejen usnadnilo a zpříjemnilo lexikografickou práci, ale také umožnilo získání přesnějších a na autentickém jazykovém materiálu založených výsledků. Jazykový korpus se zdá být pro lexikografa v současné době nepostradatelným nástrojem a zdrojem dat (např. Blatná – Čermák 1995).

Pro tvorbu dvojjazyčných překladových slovníků jsou v současnosti bezesporu nejvhodnější paralelní korpusy. Paralelní korpusy<sup>5</sup> jsou korpusy dvou nebo více jazyků vytvářené z překladů, obsahující tedy vždy jednak texty originální, jednak jejich jinojazyčné mutace (Kocek – Kopřivová – Kučera 2000, 9). Paralelní korpusy umožňují vyhledávání ekvivalentů přímo v prostředí jejich užití v daném jazyce, zapojených v kontextu, a tím dávají možnost získat přesnější obraz o kontextové zapojenosti a kontextových významech vyhledávaného slova. Na FF UK v Praze je v současné době realizován projekt Intercorp, který má umožnit vznik paralelních korpusů pro většinu jazyků vyučovaných na FF UK. Pro malé jazyky, mezi které můžeme řadit češtinu i

---

<sup>5</sup> Někteří korpusoví lingvisté považují termín paralelní korpus za zastřešující pojem pro různé druhy korpusů. Například Altenberg a Granger (2002, 7-8) vyčleňují termíny srovnávací korpus (comparable corpora) – soubory podobně tematicky zaměřených původních textů v různých jazycích a překladový korpus (translation corpora) – totožné texty v různých jazycích (originál a jeho překlady). Problémem překladových korpusů je to, že překlady nemusí být přirozené texty daného jazyka, při překladu často dochází k ovlivňování podoby textu jazykem originálního textu.

finštinu, je hlavním problémem tvorby paralelního korpusu nedostatečné množství použitelného materiálu, předpokladem realizace projektu je možnost využití odpovídajících textů v obou jazycích. U malých jazyků je právě nedostatek vhodných textů, společně s problematikou autorských práv použitelných textů hlavní příčinou obtíží při tvorbě paralelních korpusů. Dalším problémem je nevyváženost a nedostatečná reprezentativnost paralelních korpusů, jejich rozsah a často i omezené tematické zaměření textů v nich obsažených zdaleka nedosahují rozsahu a reprezentativnosti velkých jednojazyčných korpusů (Altenberg – Granger 2002, 34). Proto jedním z dalších možných zdrojů materiálu použitelného pro komparativní výzkum malých jazyků zůstávají jejich jednojazyčné korpusy. Jejich nevýhodou je sice to, že jsou oddělené a nedají se v nich tedy porovnávat odpovídající texty, svou velikostí jsou ale neporovnatelně reprezentativnějšími zdroji jazykového materiálu než stále jen kusé paralelní korpusy.

## 2 Teoretický rámec

### 2.1 Jazykový korpus

Korpusová lingvistika je disciplína lingvistiky zkoumající jazyk pomocí elektronických jazykových korpusů a zabývající se i výstavbou těchto korpusů, jejich zpracováním a příslušnou metodologií (Kocek – Kopřivová – Kučera 2000, 4). Je tedy disciplínou spojující lingvistiku s dalšími obory – matematikou, informatikou, statistikou, či počítačovou technikou. Základním prostředkem výzkumu je pro korpusovou lingvistiku jazykový korpus. Je to strukturovaný unifikovaný a často též označovaný soubor jazykových dat, který je elektronicky uložený i zpracováváný (např. Kocek – Kopřivová – Kučera 2000; Blatná – Čermák 1995). Jak ale uvádí např. Kennedy (1998), či Tognini-Bonelli (2001), definice konkrétního korpusu závisí také na jeho struktuře a účelu, korpus má být charakteristickým obrazem jazyka, texty v něm obsažené mají zachycovat zákonitosti jazyka, či jeho variant.

Jedním ze základních východisek korpusové lingvistiky je zkoumání jazykových jevů v jejich přirozeném prostředí, v jejich situačním kontextu. Pojem situační kontext (context of situation) zavedl do lingvistiky J.R. Firth (1957) a chápal ho jako centrální koncept celé sémantiky. Firth (1957, 10) rozlišil tři základní složky slovního významu – centrální (denotační), neboli věcný význam, expresivní (pocitový) význam a význam kontextový. Komplexní význam slova obsahuje všechny tyto tři složky a je plně realizován ve svém konkrétním užití v kontextu situace. Kontext je přitom pro Firtha široký pojem zahrnující jazykové i nejazykové kontexty mluvčího, kontexty historické, sociální a kulturní.

Jazykový korpus nabízí možnost zkoumání lexikálních jednotek v jejich konkrétním užití, v situačním textovém kontextu. Při práci s jazykovým korpusem máme možnost pracovat s textem vertikálně i horizontálně – konkordanční programy řadí jednotlivé výskyty například do sloupců, v nichž můžeme sledovat zkoumaný jev v jeho různých kontextech výskytu, samotný sledovaný jev je přitom zobrazen jako zarovnaný sloupec – vertikálně, kontext jednotlivých výskytů pak horizontálně po obou stranách vyhledaného jevu.

## **2.2 Zkoumání jazyka jako produktu komunikace**

Korpusová lingvistika zkoumá přirozené, autentické texty, texty vytvořené skutečnými mluvčími konkrétních jazyků (Sinclair 1991, 5). Autentičnost, a jak uvádějí např. McEnery, Xiao a Tono (2006, 80) v mnoha autentických textech realizovaná intuice velkého množství mluvčích – autorů korpusových textů, jsou hlavní opěrné body pro korpusová data. Korpusová lingvistika zkoumá tedy jazykovou realizaci, v Saussurově pojetí parole a v pojetí Chomského performanci (Tognini-Bonelli 2001; Kennedy 1998)<sup>6</sup>. Zkoumání jazykové performance je zároveň jednou z cest k poznání jazykového systému (langue/kompetence), sledováním jazykové produkce mluvčích můžeme bez použití kognitivních, či generativních modelů získat informace o fungování jazyka jako abstraktního systému (Kennedy 1998).

Pro možnosti zkoumání reálného stavu jazyka je nutné, aby data obsažená v korpusu odrážela skutečnou situaci v jazyce – korpus musí být reprezentativní (Kennedy 1998). Reprezentativnost je důležitým faktorem při výstavbě textového korpusu. Faktory souvisejícími úzce s reprezentativností korpusu je rozsah zkoumaných

---

<sup>6</sup> Jsme si ovšem vědomi toho, že Saussureho a Chomského pojetí se zcela nepřekrývají.

dat, tedy objem korpusu, autentičnost textů, vyváženost, balancovanost korpusu, tedy jeho obsah a v neposlední řadě také možnost zobecnění výsledků korpusových analýz na jazyk jako celek.

Korpus musí obsahovat dostatek materiálu, aby bylo na jeho základě možné vyvozovat adekvátní a opodstatněné závěry. Velikost korpusu souvisí s jeho účelem a také se zaměřením konkrétního výzkumu. Pokud se rozhodneme zkoumat méně frekventované jevy, potřebujeme velký korpus (v současné době velké korpusy obsahují řádově stamilióny slov, ale se stoupajícími paměťovými kapacitami počítačů a možnostmi využití dalších forem textů velikost korpusů rychle roste). Při studiu frekventovaných jevů nastává naopak problém s analyzovatelností získaných dat, z velkého množství materiálu je pak třeba vybrat jen reprezentativní vzorek (například využitím funkce náhodného výběru).

Protože je stále ještě jednodušší zaznamenávat a zpracovávat psané texty než texty mluvené, i když je jejich produkce v porovnání s produkcí textů mluvených ve velké menšině, jazykové korpusy vznikají nejčastěji právě jako soubory textů psaných. Pro vyšší podíl psaných textů v korpusech hovoří i fakt, že mají větší vliv na uživatele jazyka – psaný text si většinou přečte více lidí, než kolik si jich vyslechne text mluvený (Kennedy 1998). Mluvený jazyk ale právě kvůli jeho velké kvantitativní převaze nelze opomíjet. Současné skutečně reprezentativní korpusy obsahují i určitý podíl textů mluvených.

Neméně důležitým měřítkem reprezentativnosti korpusu je obsah korpusu, tedy typy a původ textů v něm obsažených. Korpus by měl být pestrý a zároveň by měl odrážet produkci různých typů textů, jejich recepci u čtenářů a také jejich jazykovou relevanci. Rozhodující je zastoupení základních literárních druhů tedy beletrie,

publicistiky a odborné literatury. Například Český národní korpus obsahuje pestrou škálu textů, které byly vybrány na základě statistických výzkumů zastoupenosti různých textových typů a žánrů v psaném jazyce<sup>7</sup>. Texty jsou roztrženy do různých kategorií podle témat a oborů, k nimž náležejí. Zastoupení jednotlivých druhů textů v ČNK uvádí tabulka č. 1. (Podrobně se výstavbou ČNK zabývají ve svém článku Králík a Šulc (2005)).

	SYN2005	SYN2000
Beletrie	40%	15%
Odborná literatura	27%	25%
Publicistika	33%	60%

**Tabulka č. 1** ukazuje zastoupení jednotlivých druhů textů v ČNK. První sloupec uvádí zastoupení jednotlivých druhů textů v subkorpusu ČNK SYN2005, druhý sloupec situaci v SYN2000.

### 2.3 Kolokace

V jazykových textech se slova nevyskytují osamoceně, ale sdružují se do vyšších celků, skupin slov, vět a souvětí. Zkoumání zákonitostí spojování slov, jejich vzájemného souvškytu a s tím souvšející aktualizace jejich významů je důležitý faktor, který je třeba brát na vědomí při vytváření slovníkových lemmat. Jak se význam slova obměňuje, když se slovo spojuje s jinými slovy, která spojení jsou volná a zaměnitelná a která jsou pevná a neoddělitelná?

Korpusová lingvistika umožnila svými postupy zkoumání spojitelnosti slov a jejich chování v jazyce. Jedním z hlavních pojmů současné korpusové lingvistiky se stal pojem kolokace – kombinace jazykových prvků lexikální povahy (Encyklopedický

<sup>7</sup> Více take na www stránkách <http://ucnk.ff.cuni.cz/korpus/korpus>



slovník češtiny 2002). Kolokace, jak uvádí Čermák (Čermák – Šulc 2006), v češtině nahrazuje související a již nevyhovující pojem slovní spojení. Pojem kolokace zavedli do jazykovědy Palmer s Hornbym ve 30. letech 20. století a ve svých pracích ho oživil J. R. Firth (1957). V souvislosti s korpusovou lingvistikou jej později rozvedli J. Sinclair, M. A. K. Halliday a další.

Firth chápal kolokaci jako charakteristický, habituální spoluvýskyt jazykových jednotek vázaný na situační kontext (Firth 1957, 194). Spojování slov, nebo pro lepší orientaci lexikálních jednotek, jak je formuluje Sinclair (1998, 23), funguje podle Hallidaye (1966, 152) na bázi gramatických vztahů, do nichž jsou konkrétní formy slov – lexikální jednotky – obsazovány jako prvky lexikálního paradigmatu jazyka. Ve větě jsou gramatické pozice obsazovány vhodnými konkrétními tvary lexikálně-sémantických jednotek a vzniká tak systém gramaticko-sémantických kolokací. Pokud se ale stejné kolokace vyskytují na různých místech v různých větách, je jejich stavba – souvýskyt jejich členů – motivována jak vnitřně, na základě gramatických pravidel (forma) a lexikálně-sémantických vztahů (obsah), tak zároveň zvnějšku obsazením konkrétních pozic v dané větě. V takovémto případě pak kolokace fungují jako celistvé jednotky, které jsou sice složeny z více segmentů, které jsou ale zároveň nerozložitelné v dané pozici a kontextu v konkrétním obsazení ve větě.

## 2.4 Metody *corpus-based* a *corpus-driven*

Korpusová lingvistika rozlišuje dva hlavní přístupy pro práci s daty získanými z jazykových korpusů – metody *corpus-based* a *corpus-driven* (Tognini-Bonelli 2001). První z metod – *corpus-based* (korpus jako prostředek ověřování dat) – využívá jazykové korpusy pro testování a ověřování předdefinovaných nebo již existujících

teorií. Druhá metoda – *corpus-driven* (korpus jako zdroj dat) – využívá naopak korpus jako podklad pro vytváření jazykových teorií, korpusová data jsou pro ni materiálem pro teoretické formulace.

Jak uvádí Tognini-Bonelli (2001, 66), lingvisté, kteří své předdefinované teorie ověřují pomocí korpusových dat, většinou jazykový korpus chápou jako pouhou vedlejší metodu, která jim může pomoci jejich teorii ověřit a podpořit, ale je jen jakýmsi bonusem, nemá sílu předdefinovanou teorii ve větší míře ovlivnit. Metoda *corpus-driven* naproti tomu vychází z korpusových dat jako základního zdroje poznání jazyka, je empirickou metodou založenou na kontextuálním a funkčním přístupu k jazyku (Tognini-Bonelli 2001, 87). Dostatečně velký, reprezentativní a vyvážený korpus je schopen být autentickým vzorkem přirozeného jazyka a je schopen podat s použitím vhodných technologií informace o vzorcích, kontextech a frekvencích zkoumaných jazykových jevů v rámci parole. Metoda *corpus-driven* přináší možnost sledovat language prostřednictvím parole – každá korpusová konkordance je jedinečnou jazykovou realizací, která může být náhodná, ale také opakovaná a tedy typická.

Pro sledování fungování lexémů v jazyce je důležité mít možnost zkoumat dané lexémy v kontextech jejich užití. Kolokační a kontextuální zapojení lexémů velice často modifikuje jejich význam. Tyto modifikace není možné zjistit ani zachytit zkoumáním lexémů bez jejich zapojení do kontextu. Jazykový korpus umožňuje zkoumat lexémy v širokém a volitelném kontextu a napomáhá tak zmapovat proměny jejich významů ve spojení s různým okolím.

Na závěr tohoto oddílu je ovšem nutno podotknout, že obě tyto metody jsou v praxi spíše tendencemi, je těžké hledat hranici mezi pouhým ověřováním jazykových dat v korpusu a skutečným využitím dat z korpusu získaných. Proto ani v případě této

práce netvrdíme, že náš postup je čisté využití metody *corpus-driven*, jedním z našich cílů je totiž srovnání dat získaných z korpusu s daty ze starších nekorpusových slovníkových prací. Jazykovědná zkoumání musejí totiž využívat pestrých možností jazykovědy a ověřovat materiály z různých zdrojů.

## 3 Metodologie a data

### 3.1 Substantiva vyjadřující čas

Jako materiál pro svou korpusovou analýzu jsme si zvolili skupinu substantiv vyjadřujících čas. K této volbě nás vedla potřeba získat pro analýzu vhodnou skupinu jednotek, které by tvořily homogenní celek, mezi nimiž by byly i jiné, než sémantické, či gramatické souvislosti, a které by i svým počtem odpovídaly potřebám naší analýzy.

Námi zvolená skupina substantiv obsahuje celkem 8 lexémů. Po sémantické stránce se jedná o slova vyjadřující různě dlouhé časové etapy a to neurčitě a proměnlivě dlouhé (*čas/aika*), relativně ustálené, i když různými kulturami a tradicemi různě vymezené (*rok/vuosi*, *měsíc/kuukausi*, *týden/viikko*, *den/päivä*) i přesně definované (*hodina/tunti*, *minuta/minuutti*, *sekunda/vteřina/sekunti*). Kromě gramatických a sémantických souvislostí je pro tuto skupinu typické i to, že v ní obsažené lexémy patří do základní slovní zásoby. Jedná se o stará slova, patřící do původní slovní zásoby, případně o slova přejatá v souvislosti s rozvojem techniky a přesnějším počítáním času. Důležitým hlediskem výběru byl také fakt, že se jedná o slova frekventovaná. Podle Frekvenčního slovníku finštiny (Saukkonen a kol., 1979) všechna tato substantiva patří do první tisícovky nejčastěji se vyskytujících slov ve finštině. Substantiva *aika*, *vuosi* a *päivä* patří do stovky nejfrekventovanějších slov, substantivum *aika* je navíc nejfrekventovanějším substantivem vůbec, následuje ho *vuosi*, substantivum *päivä* je sedmé nejčastější substantivum ve finštině (viz tabulka 2).

	<i>Frekvenční pořadí</i>	<i>Počet výskytů (F)</i>	<i>F/408301x10000 / ‰</i>	<i>F/beletrie</i>	<i>F/rádio</i>	<i>F/publ.</i>	<i>F/odb.</i>
<b>aika*</b>	20	1347	32,99	32	22	18	24
<b>vuosi</b>	23	1232	30,17	113	30	19	20
<b>kuukausi</b>	493	115	2,82	1069	368	411	603
<b>viikko</b>	308	174	4,26	445	137	294	677
<b>päivä</b>	51	693	16,97	38	40	58	64
<b>tunti</b>	350	157	3,85	307	240	463	421
<b>minuutti</b>	754	76	1,86	935	328	786	1526
<b>sekunti</b>	1249	45	1,1		583	874	2842

**Tabulka č. 2 ukazuje frekvenci substantiv vyjadřujících čas ve finštině** (Saukkonen, Pauli; Haipus, Marjatta; Niemikorpi, Antero; Sulkala, Helena: Suomen kielen taajuussanasto.WSOY. Porvoo, Helsinki, Juva 1979.). Frekvence byly vypočteny na základě materiálu, který obsahoval 408 301 slov. První sloupec uvádí pořadí frekvence ve frekvenční tabulce, druhý sloupec absolutní frekvenci výskytu. Ve třetím sloupci je vypočtená relativní frekvence vztažená k celkovému počtu slov. Poslední čtyři sloupce uvádějí přepočtené pořadí ve frekvenční tabulce pro beletrii, mluvené texty, publicistiku a odbornou literaturu.

\* pouze pro substantivum aika, adverbium uvedeno zvlášť

### 3.2 Kielipankki

Základním prostředkem získávání dat o zkoumaných lexikálních jednotkách ve finštině nám byl korpus finského jazyka Kielipankki. Kielipankki<sup>8</sup> je služba poskytovaná Finským informačním centrem pro vědu (Tieteen tietotekniikan keskus, Finnish IT Center for Science, CSC) skládající se z několika na sobě nezávislých textových korpusů, korpusových manažerů a programů. Textová část Kielipankki

<sup>8</sup> Kielipankki – „jazyková banka“, angl. The Language Bank

obsahuje korpusy mluvených a psaných textů několika jazyků (kromě finštiny také angličtiny, švédštiny, němčiny, francouzštiny, svahilštiny).

Korpus finských textů (Suomen tekstikokoelma, Finnish Text Collection, fct) je součástí Kielipankki. Je to soubor psaných finskojazyčných textů z 90. let 20. století. Jedná se z velké většiny o texty publicistické, korpus obsahuje celé ročníky některých významných finských deníků. Korpus finských textů byl vytvořen na základě spolupráce institucí Kotimaisten kielten tutkimuskeskus, Helsingin yliopiston Yleisen kielitieteen laitos a Joensuun yliopiston Vieraiden kielten laitos<sup>9</sup>. Celkem tento korpus obsahuje 97 subsouborů, které jsou složeny z 645 274 dokumentů. Velikost korpusu je 180 milionů tokenů<sup>10</sup>. Grönroos (2000) uvádí, že v září 2000 obsahovala tehdejší Suomen kielen tekstipankki 175 milionů slov, z nichž přibližně 1/3 byla morfosyntakticky otaggována<sup>11</sup>. V současné době je otaggováných téměř 170 milionů slov, samotná materiálová základna se ale od roku 2000 příliš nerozrostla.

Práce se textovým korpusem finského jazyka probíhá prostřednictvím [www](http://www.lemmikki.org/) rozhraní WWW-Lemmie 2.0. WWW Lemmie 2.0 je dynamický internetový prohlížeč, s jehož pomocí lze pracovat s texty Suomen tekstikokoelma on-line. Je internetovou verzí Lemmie API 2.0, objektově založeného programového rozhraní pro práci s lexikální databází Kielipankki, vytvořeného v programovacím jazyku Perl. Obě verze jsou propojené a to umožňuje jejich uživatelům kombinovat práci s oběma verzemi a ověřovat a potvrzovat výsledky výzkumů jak v online rozhraní WWW Lemmie 2.0, tak

---

<sup>9</sup> Centrum výzkumu domácích jazyků, Katedra obecné lingvistiky Univerzity v Helsinkách, Katedra cizích jazyků Univerzity v Joensuu

<sup>10</sup> Korpusová lingvistika rozlišuje několik termínů týkajících se výskytů slov/slovních tvarů v korpusu. Termín *token* označuje každý konkrétní výskyt slova v korpusu, počet tokenů je tedy celkový počet slov v korpusu. Termín *type* naproti tomu označuje jednotlivé konkrétní tvary/formy slov vyskytující se v korpusu nezávisle na počtu jejich výskytů (Oakes 1998, 256).

<sup>11</sup> tag – morfologická značka. Většina současných jazykových korpusů je otaggována – každý výskyt slova v korpusu je označen morfologickými značkami (např. slovní druh, morfologické kategorie apod.), které zjednodušují vyhledávání a umožňují různé druhy vyhledávání.

přímo v programu Lemmie API 2.0. Přístup do WWW Lemmie 2.0 je možný přes vědecké uživatelské rozhraní CSC<sup>12</sup>, pro jeho užívání je třeba získat přihlašovací jméno a heslo na základě žádosti o přidělení uživatelských práv. Lemmie 2.0 umožňuje práci s vybranými finsko- a švédskojazyčnými texty obsaženými v Kielipankki. Jedná se převážně o texty publicistické, součástí přístupné textové sbírky jsou ale i texty z produkce nakladatelství Otava a Werner Söderström.

Pro naši práci byly relevantní pouze texty finskojazyčné. Publicistické texty jsou v Kielipankki reprezentovány celými ročníky vybraných finských deníků (Aamulehti 1995, 1999; Demari 1995, 1997, 1998, 1999, 2000; Hämeen sanomat 1999, 2000; Helsingin sanomat 1995; Hyvinkään sanomat 1994, 1997; Iltalehti 1996; Kaleva 1998-1999; Kangasalan sanomat; Karjalainen 1991-1995, 1997-1999; Keskisuomalainen 1999; Turun sanomat 1998, 1999), texty ze tří ročníků magazínu Tekniikan maailma (1995-1997) a několika texty z produkce výše jmenovaných nakladatelství. Celkem finskojazyčná část textů obsahuje 167 010 084 slov.

Při práci s korpusem je možné vybrat a přednastavit texty, s nimiž se bude dále pracovat. To umožňuje omezit velikost zkoumaného materiálu a do určité míry také vybrat například geografickou oblast, jejíž jazyk se rozhodneme zkoumat (vybráním textů některého z regionálních deníků). Bohužel z důvodu nevyrovnaného složení korpusu, v němž naprosto převažuje publicistika (beletrie korpus obsahuje pouze 1,6 %) je možné zkoumat téměř jen texty publicistické. Na druhou stranu i texty publicistické mohou poskytnout dosti pestrý jazykový materiál. Jak uvádí Landau (2001) velké deníky jsou ideálním zdrojem textů pro korpus, protože obsahují celou škálu tematicky pestrých textů, od zpráv z politiky, sportu a byznysu po kuchařské recepty a filmové či knižní recenze. Přesto je třeba mít stále na mysli, že psaný publicistický text je jen

---

<sup>12</sup> <https://hotpage.csc.fi/log/in.phtml>

jednou z forem jazyka, nemůžeme ho tedy považovat za reprezentativní vzorek celého jazyka.

Pro analýzu substantiv vyjadřujících čas ve finštině jsme se rozhodli omezit pracovní materiál jen na texty nejčtenějšího finského deníku Helsingin Sanomat, který je v Kielipankki zastoupen ročníkem 1995. Celkem 22 subsouborů, do nichž jsou Helsingin sanomat rozděleny, obsahuje 22 064 252 slovních tvarů. Základní informace o Helsingin sanomat a dalších velkých finských denících podává tabulka č. 3.

Název deníku	Místo vydání	Průměrný náklad	Počet čtenářů	Vychází dnů/týden Pořadí v počtu čtenářů
Aamulehti	Tampere	138 258	316 000(2007)	7/t 2.
Helsingin Sanomat	Helsinki	417 548	1 030 000	7/t 1.
Kaleva	Oulu	82 037	206 000	7/t 4.
Keskisuomalainen	Jyväskylä	75 865	181 000 (2007)	7/t 5.
Turun Sanomat	Turku	111 547	263 000	7/t 3.

**Tabulka č. 3** představuje 5 nejčtenějších finských deníků, jejichž vybrané ročníky tvoří podstatnou část Kielipankki.

### 3.3 Práce s WWW Lemmie 2.0

Práce s Lemmie 2.0 se v mnoha ohledech liší od práce s ČNK. Protože vyhledávání probíhá přes www rozhraní, komunikace se serverem je s ohledem na velikost dat, která Lemmie 2.0 zpracovává, dosti zdlouhavá. Lemmie navíc umožňuje v jednom okamžiku pracovat maximálně s 20 000 konkordančními řádky, které ukládá do mezipaměti. Frekventovanější výrazy (v případě substantiv vyjadřujících čas se jedná o vysoce frekventované výskyty) je proto nutné zkoumat postupně, často je tak třeba vygenerovat i několik dvacetitisícových (nebo odpovídající počet méně obsáhlých) seznamů konkordancí v závislosti na přednastaveném obsahu mezipaměti. Každý



seznam konkordancí<sup>13</sup> pak funguje jako samostatný výsledek, to se odráží například ve výpočtech frekvencí, které jsou počítány pouze v rámci vygenerovaného aktuálního seznamu, ale vztaženy k celkovému počtu slov v přednastaveném korpusu. V praxi proto frekvence výrazů, jejichž výskyt přesahuje maximální objem mezipaměti, ztrácí svou výpovědní hodnotu. Lemmie neobsahuje funkci náhodného výběru, který by umožnil zobrazení vzorku výskytů vyhledávaného výrazu. Vyhledané konkordance jsou navíc řazeny podle svého výskytu v korpusu, takže jednotlivé dvacetitisícové vzorky se od sebe mohou výrazně lišit. Tento hendikep jsme se rozhodli řešit výraznějším využitím druhotných třídících programů, tedy programů Microsoft Excel a Microsoft Access.

Lemmie zobrazuje vyhledávané výrazy jako KWIC (Key Word in Context)<sup>14</sup> s nastavitelným rozsahem kontextu vlevo i vpravo od klíčového slova v rozmezí 0 slov až hranice odstavce. Klíčové slovo a slova v kontextu lze zobrazit v několika možných formátech (jako word, lemma, tag atd.). Lemmie obsahuje také základní statistické funkce – frekvenční a kolokační tabulky.

### 3.4 Český národní korpus

Český národní korpus je projekt, jehož cílem je budování korpusů českého jazyka v jeho různých podobách. Byl založen roku 1994 a v současnosti se skládá z několika podkorpusů v různých formách. Jeho dvě základní části jsou část synchronní a diachronní, synchronní část pak obsahuje mluvené korpusy (Pražský mluvený korpus, Brněnský mluvený korpus, ORAL 2006) a psané korpusy reprezentované především

---

<sup>13</sup> Konkordance je konkrétní výskyt zkoumaného jevu v korpusu zobrazený většinou v minimálním kontextu (např. věty, odstavce apod.)

<sup>14</sup> Klíčové slovo v kontextu

korpusy SYN2000, SYN2005 a SYN2006PUB, obsahuje ale např. i menší korpusy korespondence nebo korpusově zpracovaný román George Orwella 1984. Korpusy SYN2000 a SYN2005 obsahují každý 100 milionů textových slov, SYN2006PUB celkem 300 milionů textových slov. Velikost synchronní části ČNK je tak v současné době celkem více než půl miliardy tokenů.

Práce s texty ČNK probíhá prostřednictvím korpusového manažeru Bonito, novinkou je možnost využití on-line přístupu přes server World Sketch Engine bez nutnosti stahování programu Bonito. Obě verze pracují se stejnými korpusovými daty, minimálně se liší jen některé jejich funkce.

Český národní korpus jsme použili jako základní materiál pro získání informací o jazykovém obrazu zkoumaných lexémů v češtině. Při své práci s ČNK jsme používali především korpusový manažer Bonito, pro některé frekvenční výpočty i on-line verzi ČNK. Ze subkorpusů ČNK byl hlavním zdrojem našich zkoumání korpus SYN2000, který je svým složením i dobou vzniku bližší Kielipankki, než například SYN2005. SYN2000 obsahuje 60 procent publicistických textů, které pocházejí především z 90. let 20. století, časově a zčásti tedy i zpracovávanými tématy se jeho publicistické texty shodují s publicistickými texty v Kielipankki. Svá zkoumání jsme ale ověřovali i sondami do subkorpusů SYN2005 a z důvodu možností srovnání publicistických textů (vzhledem k výrazné převaze publicistických textů ve finském korpusu) i korpusu SYN2006PUB.

## 4. Korpusové analýzy

### 4.1 Statistiky

#### 4.1.1 Statistické údaje

Jedním z možných způsobů zkoumání korpusového materiálu je statistický výzkum. Základními statistickými údaji, které lze získat z korpusu, jsou informace o frekvencích výskytu zkoumaných jevů v jazyce. Absolutní frekvence, tedy celkový počet jednotek zkoumaného jevu nalezených ve zkoumaném vzorku jazyka, v našem případě v korpusu, je číslo, které nám může podat informace o hojnosti výskytu daného jevu (jde-li o číslo vysoké, jde o jev hojný, jde-li o číslo malé, jev řídký), ale až vztahením k celkovému počtu všech jazykových jednotek ve zkoumaném vzorku jazyka, tedy výpočtem relativní frekvence zkoumaného jevu, získáme číslo směrodatné a analyzovatelné ve vztahu k podobným nebo příbuzným jazykovým jednotkám, popřípadě dalším členům skupiny zkoumaných jevů.

Jak jsme se již zmínili v teoretické části práce, skupina substantiv vyjadřujících čas je statisticky významnou skupinou substantiv řadících se k nejfrekventovanějším slovům v jazyce. Jejich zevrubnější zkoumání proto může díky možnosti studia obsáhlých korpusových materiálů přinést výsledky zajímavé pro jejich popis ve slovnících. Údaje o absolutní a relativní frekvenci jsou z korpusu celkem snadno získatelné, a proto jsme k jejich analýze přistoupili v prvním kroku. Tabulka č. 4 uvádí srovnání frekvencí substantiv vyjadřujících čas ve Frekvenčním slovníku finštiny (Saukkonen a kol., 1979) a v námi zkoumané části Kielipankki.

Ačkoli se materiály, z nichž byly získány informace o frekvencích výskytu substantiv vyjadřujících čas ve Frekvenčním slovníku finštiny a Kielipankki znatelně odlišují (časově oba jazykové vzorky dělí více než dvacet let, frekvenční slovník

pracoval s materiálem obsahujícím jen několik stovek tisíc slov, námi použitá část Kielipankki zase obsahuje jen publicistické texty), je jistě zajímavé, že údaje o frekvenci časových substantiv v obou vzorcích jazyka se nijak příliš nerozcházejí. Největší rozdíl je na první pohled zaznamenanatelný u substantiva *aika*, ve frekvenčním slovníku finštiny je jeho frekvence více než třikrát větší. Příčinu tohoto jevu lze zřejmě bez bližších zkoumání načrtnout jen v dohadách, vyšší frekvenci substantiva *vuosi* v publicistických textech lze naopak jasně vysvětlit častým užíváním přesných časových odkazů na roky (*vuonna 2000* – roku 2000, *vuodesta 2000* – od roku 2000) a častějším užíváním přesných časových údajů v textech zpravodajských a publicistických vůbec. Hojné užívání přesných časových údajů v textech zpravodajských a sportovně zpravodajských dokazuje také zvýšená frekvence substantiv *minuutti* a *sekunti*.

Tabulky č. 4 a 5 ukazují situaci frekvencí výskytu substantiv vyjadřujících čas ve finštině a češtině (podle Frekvenčního slovníku finštiny a Kielipankki (tabulka č. 4), a subkorpusů ČNK – SYN2000 a SYN2006PUB (tabulka č.5)).

Lexém	Abs. Kielipankki	Rel. Kielipankki	Rel.Frekv.slovník
aika	21746	9,86	32,99
vuosi	93338	42,30	30,17
kuu(kausi)	2135(7812)	0,97(3,54)	2,82
viikko	17548	7,95	4,26
päivä	20505	9,29	16,97
tunti	6799	3,08	3,85
minuutti	5516	2,5	1,86
sekunti	2841	1,29	1,1

**Tabulka č. 4** uvádí srovnání frekvencí substantiv vyjadřujících čas ve Frekvenčním slovníku finštiny a Kielipankki. Pro výpočet relativní frekvence jsme pro možnost přehlednějšího srovnání použili vzorec užitý ve Frekvenčním slovníku finštiny (Saukkonen a kol., 1979)  $F(a)/\square(\text{celk}).10000$ . U substantiva *kuu/kuukausi* – měsíc uvádíme statistiky pro obě substantiva, více o rozdílu mezi těmito dvěma lexémy v kapitole 4.1.2.

Lexém	Abs.SYN2000	Rel.SYN2000	Abs.SYN2006PUB	Rel.SYN2006PUB
čas	62174	4,82	179855	4,98
rok	406834	31,52	1538733	42,60
měsíc	44850	3,47	164536	4,56
týden	49417	3,83	205500	5,69
den	110289	8,54	361195	10,00
hodina*	77122	3,44	321739	8,91*
minuta	32639	2,53	134550	3,73
sekunda	28189	2,18	10827	0,30
vteřina	4733	0,37	17322	0,48

**Tabulka č. 5** uvádí srovnání frekvencí substantiv času v korpusech SYN2000 a SYN2006PUB. První a třetí sloupec uvádí absolutní frekvenci substantiv času ve zkoumaných korpusech, druhý a čtvrtý sloupec frekvence relativní, přepočítané opět podle modelu užitého ve Frekvenčním slovníku finštiny (Saukkonen a kol., 1979).

\* - u substantiva *hodina* v korpusu SYN2006PUB v lemmatu zahrnut i význam *hodiny/kello*.

Výše uvedené tabulky ukazují až překvapivé shody frekvencí jednotlivých substantiv času v obou jazycích. Vysoké frekvence dosahuje v obou jazycích substantivum *vuosi/rok*, v Kielipankki a žánrovým obsahem textů příbuzném korpusu SYN2006PUB si dokonce odpovídají i číselné hodnoty relativní frekvence. Největší rozdíl je naopak u substantiva *aika/čas*, zdá se, že toto v porovnání s ostatními abstraktnější substantivum se v obou jazycích chová poněkud rozdílně. Opět je ovšem třeba zdůraznit to, že i frekvence tohoto substantiva ve zkoumaných korpusových textech obou jazyků jsou si bližší, než například frekvenční informace získaná z Kielipankki a Frekvenčního slovníku finštiny.

Frekvenční statistické informace ukazují, že zkoumaná slova se vyskytují v obou jazycích zhruba stejně hojně. Důvodem velmi podobných frekvencí je u této skupiny slov určitě to, že jako časové jednotky se i v různých jazycích užívají ve stejných kontextech a tedy i zhruba stejně často. Dalším důvodem je i to, že substantiva času jsou významově celkem ustálenou skupinou lexémů a v obou jazycích jsou jejich sémantická

pole zhruba stejně obsáhlá (výjimkami jsou slova *aika* – ve finštině existuje kromě substantiva *aika/čas, období* také příslovce míry *aika* ve významu „celkem“, „dost“, které jsme ale při zpracovávání statistik do frekvencí nezahrnuly, v češtině je podobným případem slovo *hodina*, v korpusu je pod toto lemma zahrnuto i slovo *hodiny/kello*, ve statistikách ze SYN2000 uvádíme ale frekvenci jen pro význam časového substantiva).

#### 4.1.2 Pádová distribuce substantiv času ve finštině

Spojování a tedy i souvšsky slov v jazyce se řídí gramatickými pravidly, jejichž dvěma hlavními principy jsou princip paradigmatický a syntagmatický. Slova nelze spojovat zcela libovolně, pro to, aby slova mohla tvořit vyšší jazykové celky, musí být splněny určité podmínky. Spojení slov musí mít samozřejmě určitý smysl, význam. Význam je vytvářen kombinací významů jednotlivých komponentů slovního spojení, aktuální situací užití, kontextem, užitím určitých gramatických významů. Spojení lexikálních jednotek je po gramatické stránce kombinací principů syntagmatického a paradigmatického. Syntagmatické spojení, tedy vztah mezi komponenty spojení jazykových jednotek, vzniká spojením slovních tvarů vybraných z určité skupiny možností – z paradigmatu (v případě substantiv z paradigmatu tvarů daného lexému). Tvarová distribuce jednotlivých lexémů závisí na povaze konkrétního jazyka, ale také na tom, do jakých sémantických a tedy i gramatických vztahů jednotlivé lexémy vstupují.

Frekvenční distribuce pádů u námi sledovaných substantiv času ve finštině ukazují, že obě výše zmíněné tendence jsou z frekvenčních tabulek distribuce pádů lehce vysledovatelné. Substantiva času (jako jednotky času) vstupují velice často do spojení s číslovkami. Ve finštině se substantivum a adjektivum ve spojení s číslovkou

užívá v partitivním tvaru singuláru. Partitiv je také proto pád s jednou z nejvyšších frekvencí, Iso suomen kielioppi (2004) uvádí, že je třetím nejfrekventovanějším pádem ve finštině. Také u všech námi zkoumaných substantiv byl jedním ze tří nejčastějších pádů, u lexémů *tunti*, *minuutti* a *sekunti*, nejčastěji užívaných právě jako jednotek času, je partitiv jasně nejfrekventovanějším pádem.

Podobně i vysoká frekvence dalšího pádu – genitivu – souvisí s povahou finštiny jako jazyka, který využívá velmi často genitivní spojení. Vysokou frekvenci tohoto pádu způsobuje hlavně časté užití genitivního atributu (například i v případech, kdy čeština užívá adjektivního spojení (*vuoden tulot* – roční příjmy, *päivän valo* – denní světlo) a také užití genitivu ve spojení s postpozicemi, do nichž právě substantiva času vstupují velmi hojně. Proto je genitiv u námi sledovaných substantiv ve všech případech jedním z dvou nejčastějších pádů, Iso suomen kielioppi ho po nominativu řadí mezi nejčastějšími pády na druhé místo.

Vedle výše zmíněných gramatických pádů vykazují u některých substantiv času vysokou frekvenční hodnotu i další pády, v tomto případě se už ale statistická distribuce pádů u jednotlivých substantiv liší. Vysokou frekvenci vykazují u substantiv času místní pády, které ale v případě této skupiny substantiv vyjadřují časové významy. Zajímavým fenoménem časově užitých místních pádů ve finštině je to, že jejich tři možné varianty pro statické vyjádření času (časový bod nebo úsek) – inessiv, adessiv a essiv se nevyskytují pravidelně a rovnoměrně, ale každé časové substantivum preferuje při vyjadřování časového určení jiný pád. Užitím různých místních pádů u téhož substantiva se pak dají vyjádřit různé časové významy, a to opět v závislosti na konkrétním substantivu.

U substantiva *vuosi* má vysokou frekvenci nepravidelný tvar essivu (*vuonna*)<sup>15</sup>, pravidelně tvořený tvar *vuotena* měl v námi zkoumané části Kielipankki frekvenci jen něco přes 200 výskytů (oproti zhruba 20 000 u tvaru *vuonna*). Tvar *vuonna* ve spojení s číslovkou označující rok vyjadřuje české spojení „roku ...“ (*vuonna 2000* - roku 2000, v roce 2000). Také ve spojení s deiktickými výrazy, adjektivy a řadovými číslovkami se substantivum *vuosi*, podobně jako *päivä*, užívá v essivu (*tänä vuonna* – tento rok, *viime vuonna* - minulý rok, *ensi vuonna* - příští rok, *sinä päivänä* – onoho/toho dne, *eräänä kauniina päivänä* - jednoho krásného dne). Zajímavý je v tomto případě také poměr mezi tvary singuláru a plurálu. Plurální tvary essivu jsou totiž zdá se preferované ve spojení tvarů substantiv množného čísla a deiktických výrazů – kromě výše uvedených substantiv *vuosi* a *päivä* do takovýchto spojení vstupují i substantiva další, u některých z nich je pak toto plurální užití essivu dokonce častější než užití essivu singuláru (*viime vuosina* - v minulých letech, *viime kuukausina* - v minulých měsících, *edeltävinä viikkoina* - v předešlých týdnech, *viime päivinä* - v posledních dnech). Tvary essivu jsou naopak velmi řídké u substantiv vyjadřujících jednotky času (*tunti*, *minuutti*, *sekunti*).

Substantivum *viikko* preferuje ve spojení s deiktickými výrazy pád adessiv (*tällä viikolla* - tento týden, *ensi viikolla* - příští týden, *viime viikolla* – minulý týden). Prosté užití adessivu u substantiva *viikko* pak vyjadřuje české „v týdnu“ (proti „o víkendu“) (*Viikolla olemme kotona, viikonloppuna käymme Helsingissä*. Přes týden jsme doma, o víkendu pojedeme do Helsinek.) Distribuce spojení s deiktickými výrazy je pak zajímavá u substantiv *kuu* a *kuukausi*. Dvojí možné vyjádření časového úseku měsíce se odráží také v užití v různých spojeních a kontextech. Substantivum *kuu* má ve finštině

<sup>15</sup> Tvary essivu se ve finštině tvoří pravidelně koncovkou *-na/-nä* připojovanou k silnému samohláskovému kmenu (např. *talo* – dům – *talona*). Jak ale uvádí např. Lehtinen (2000, 56) několik slov si do současné finštiny zachovalo dnes už neužívané tvoření essivu od souhláskového kmenu. Příkladem je např. ustrnulý tvar slova *yksi* – jedna *ymä* (*ymä muuta* – a podobně), nebo právě nepravidelný tvar essivu substantiva *vuosi* – *vuonna*.



dva významy – prvním jeho významem je měsíc jako luna, nebeské těleso, druhým významem je vyjádření měsíce jako časového úseku, ale především v případě, kdy se jedná o deiktické určení času konkrétního měsíce (*missä kuussa?* – který měsíc/ve kterém měsíci?, *tässä kuussa* – v tomto měsíci, *viime kuussa* – minulý měsíc, *ensi kuussa* – příští měsíc, *kuun viidentenä päivänä* – pátého dne v měsíci), či vyjádření měsíce jako ohraničeného časového úseku (*kerran kuussa* – jednou měsíčně, *200 euroa kuussa* – 200 eur měsíčně). Ve spojení s číslovkou se užívá druhého substantiva *kuukausi* (12 *kuukautta* – 12 měsíců), stejně tomu je u většiny postpozických spojení (*kuukausi sitten* – před měsícem, *kuukauden päästä* – za měsíc).

Prosté užití adessivu u substantiva *päivä* (*päivällä*) vyjadřuje opozici k *yöllä* (ve dne – v noci). Českému „den“ pak ve finštině odpovídají dvě substantiva, vedle substantiva *päivä* také *vuorokausi* (den, 24 hodin). V tomto případě je pak významový rozdíl mezi spojeními *koko päivän* – celý den (od rána do večera) a *ympäri vuorokauden* – celý den, nonstop, 24 hodin. Adessiv se užívá u substantiv vyjadřujících jednotky času v časových vyjádřeních ve spojení s řadovými číslovkami – (*kolmannella minuutilla* - ve třetí minutě, *viimeisellä sekunnilla* – v poslední vteřině). U substantiva *tunti* vyjadřuje užití adessivu význam na (vyučovací) hodině (*ranskan tunnilla* – na hodině francouzštiny).

Inessiv se užívá u všech substantiv času pro vyjádření opakovaných událostí během časového úseku, Iso suomen kielioppi (2004, 1197) jeho užití v časových výrazech charakterizuje jako vyjádření trvání nebo přestávky opakujících se dějů. Částečně synonymním výrazem je ve finštině spojení s postpozicí *aikana* (*kaksi kertaa vuodessa* – dvakrát ročně, *kaksi kertaa vuoden aikana* – dvakrát během/v průběhu roku, *kolme kertaa kuukaudessa/kuussa* – třikrát měsíčně, *kolme kertaa kuukauden aikana* – třikrát během/v průběhu měsíce (v prvním případě se jedná ale vždy víceméně o

pravidelně/vícekrát se opakující děje, ve druhém případě o počet událostí během konkrétního časového úseku).

Vedle významu časových bodů a úseků vyjadřují místní pády také dynamické směřování směrem od nebo do/k časovému bodu nebo úseku. Začátek či konec časového úseku je pak vyjadřován opozicí místních pádů elativu/illativu nebo ablativu/allativu. Tyto místní pády jsou pak i v časovém významu užívány ve svých základních významech vyjadřování směru (v češtině opozice *z/do*, *od/do*). Příklady takového užití mohou být výrazy jako *vuodesta 2000 (alkaen)* -- od roku 2000 (kde je pád substantiva *vuosi* určován také vazbou s postpozicí *alkaen*), *vuoteen 2005 (asti)* do roku 2005, *ensimmäisestä minuutista viimeiseen minuuttiin* – od první do poslední minuty. Ablativ je pak u substantiva *vuosi* užíván ve významu české předložky *z* (*valokuva vuodelta 1910* – fotografie z roku 1910).

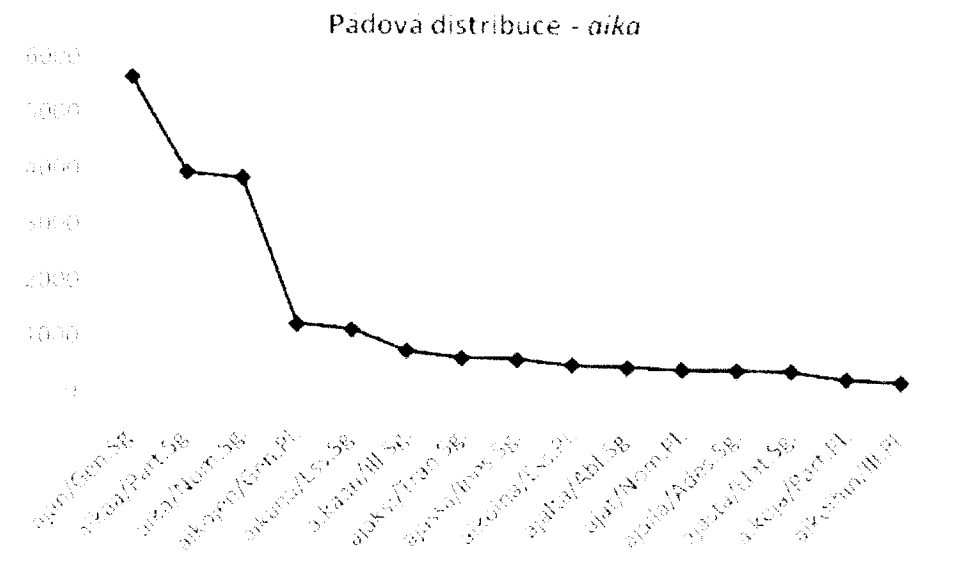
Místní pády vyjadřující směr jsou v mnoha případech užívány také nesystematicky v poloidiomatických spojeních. Iso suomen kieliooppi (2004) uvádí jako příklady užití těchto místních pádů v časových významech užití elativu při časovém určování částí dne (*tulen aamusta* – přijdu hned ráno) nebo užití ablativu při časovém určování hodin (*kahdelta* – ve dvě (hodiny), kolem druhé (hodiny)). Mnoho případů užití místních pádů v časových významech souvisí také se slovesnými vazbami či vazbami postpozic (*jäimme kuukaudeksi Pariisiin* – zůstali jsme na měsíc v Paříži, *ensi vuodesta lähtien* – od příštího roku).

Zcela zvláštním případem ve skupině substantiv času je substantivum *aika*. Na rozdíl od ostatních námi sledovaných substantiv nevyjadřuje konkrétní čas, není jednotkou času. Některé jeho tvary se ale v průběhu historického vývoje gramatizovaly a v dnešní finštině se užívají jako časové postpozice a příslovce. Jedná se především o

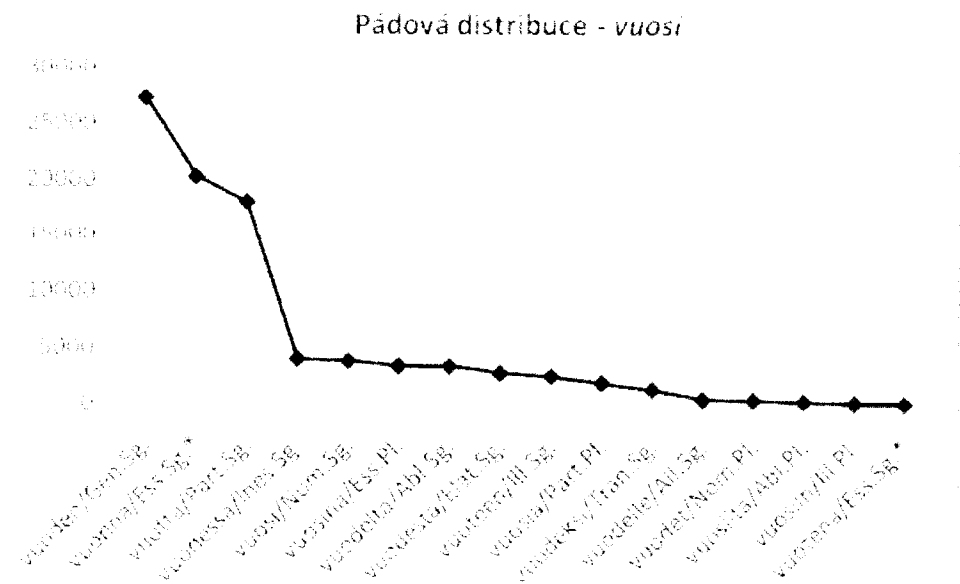
tvary/postpozice *aikana*, *ajan*, *aikaan*, které vyjadřují ve spojení s dalšími substantivy různé časové vztahy, a časová příslovce jako např. *aikaisin* - brzy, *aikaisemmin* - dříve, *aikanaan*, *aikoinaan* – „za (svých) časů“. Pádová frekvenční tabulka substantiva *aika* a také kolokační tabulky tohoto substantiva ukazují, že v Kielipankki jsou výše zmíněné postpozice otaggovány jako tvary substantiva *aika*. To sice nemůže být důkazem toho, že tyto postpozice nejsou ještě zcela oproštěny od vztahů s původním substantivem, ovšem stále velice průhledný a jasný významový vztah mezi těmito postpozicemi a substantivem *aika* dokazuje, že gramatizace těchto tvarů substantiva *aika* není ještě zcela dokončena.

Následující grafy ukazují distribuci jednotlivých tvarů substantiv času v námi zkoumané části Kielipankki. Z tabulek je zřetelně vidět, že gramatické pády (genitiv a partitiv) jsou u finských substantiv vysoce frekventované. Porovnáním tabulek lze ale také vysledovat rozdíly v distribuci místních pádů u jednotlivých substantiv času.

Grafy č. 1-8: Pádová distribuce substantiv času ve finštině

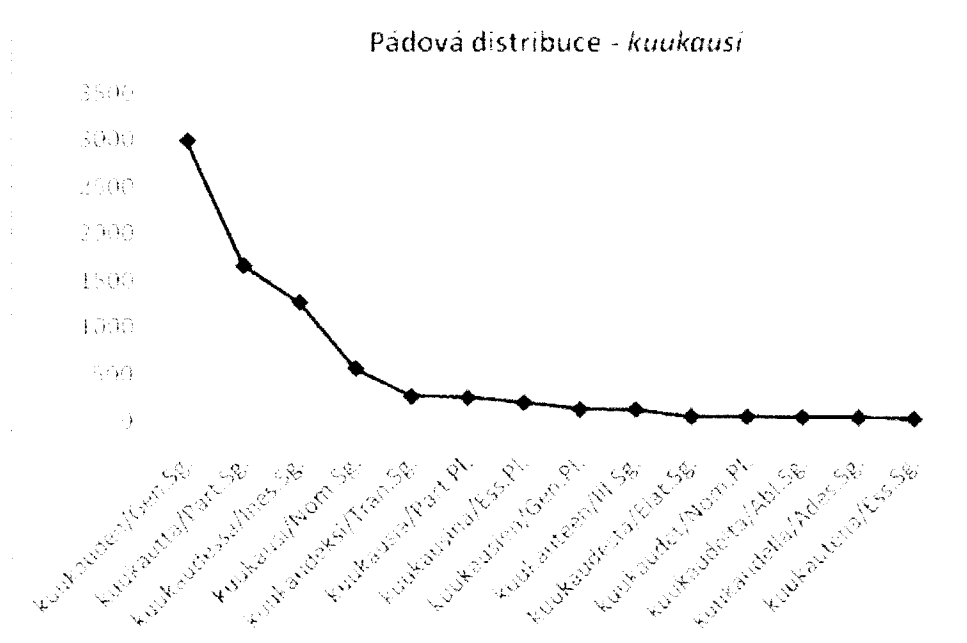


Graf č. 1 zobrazuje pádovou distribuci substantiva *aika*.

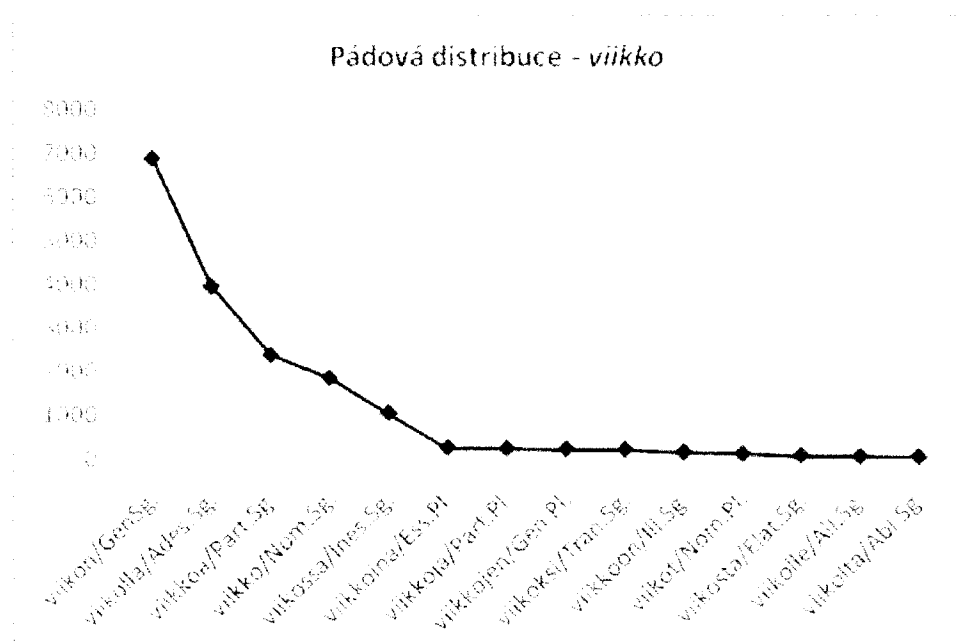


Graf č. 2 zobrazuje pádovou distribuci substantiva *vuosi*.

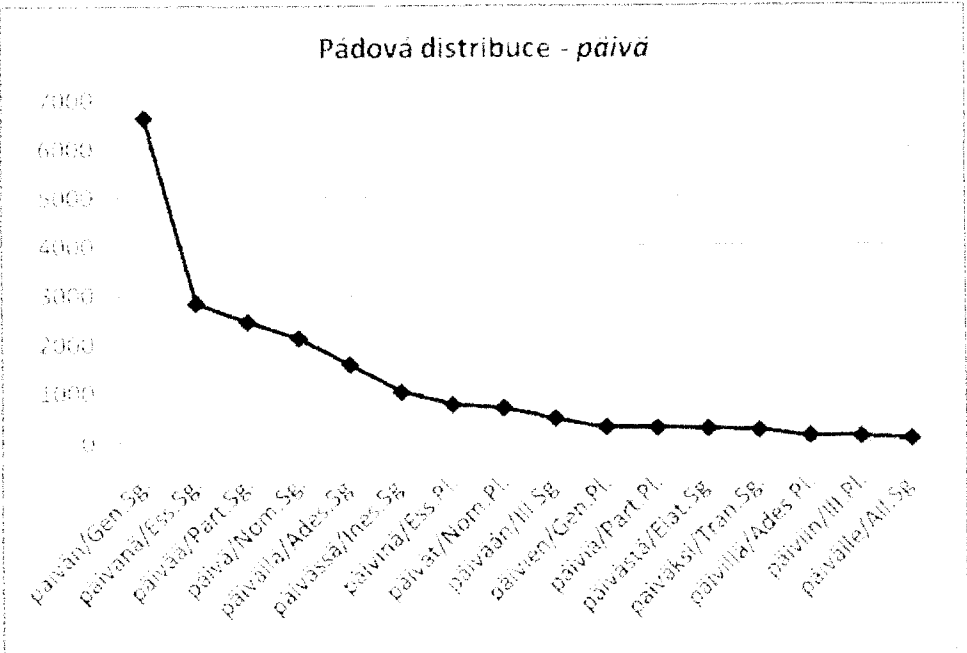
\* nepravidelný tvar essivu *vuonna* má mnohem vyšší frekvenci než pravidelně tvořený tvar *vuotena*, viz kapitola 4.1.2.



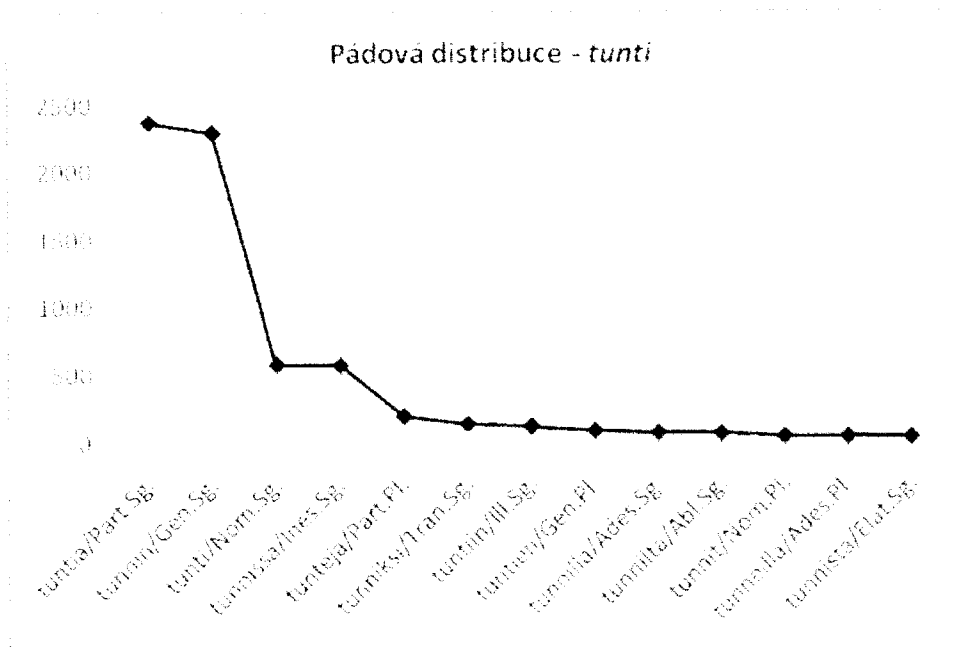
**Graf č. 3** zobrazuje pádovou distribuci substantiva *kuukausi*.



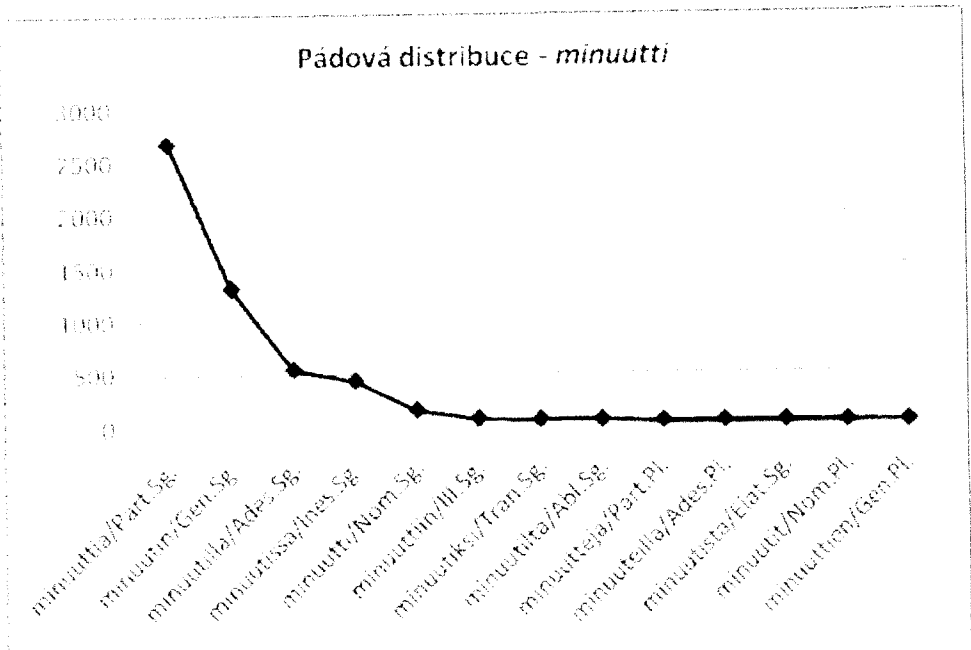
**Graf č. 4** zobrazuje pádovou distribuci substantiva *viikko*.



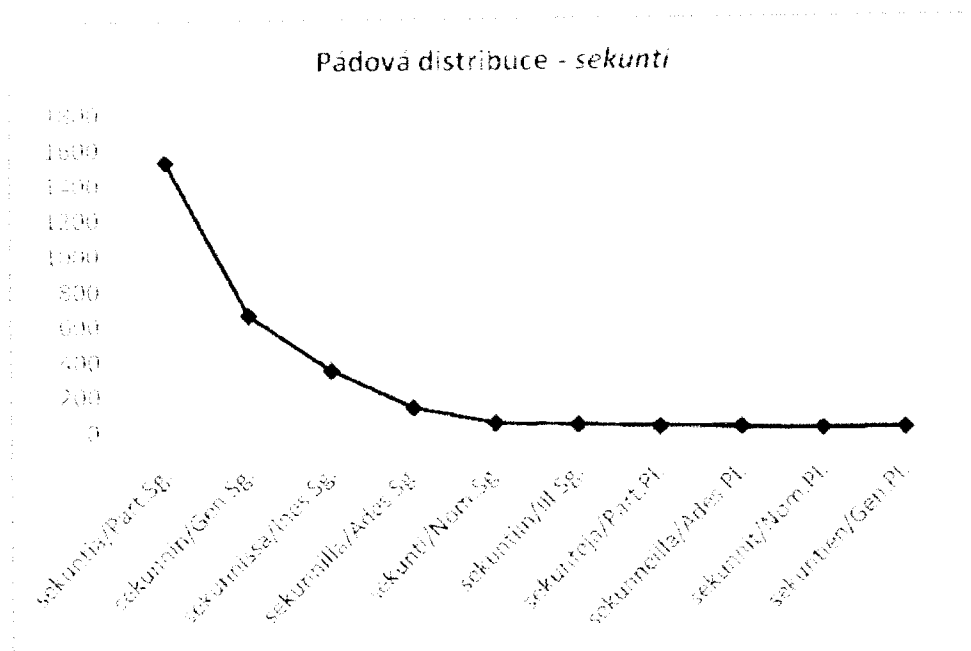
Graf č. 5 zobrazuje pádovou distribuci substantiva *päivä*.



Graf č. 6 zobrazuje pádovou distribuci substantiva *tunti*.



**Graf č. 7** zobrazuje pádovou distribuci substantiva *minuutti*.



**Graf č. 8** zobrazuje pádovou distribuci substantiva *sekunti*.

### 4.1.3 Míra asociace

Vedle absolutní a relativní frekvence je zajímavou a důležitou informací statistická pravděpodobnost souvýskytu slov, informace o míře asociace. Pro výpočet slovní kolokability existuje mnoho různých vzorců, z nichž pro naše zkoumání jsou nejvýhodnější údaje o vzájemné informaci (mutual information, MI score) a míře kontrastu (t-score). Obě tyto statistické hodnoty podávají informace o vzájemném souvýskytu dvou (nebo více) lexikálních jednotek, ale každá z nich k tomuto srovnání přistupuje jinak. MI score podává informaci o síle souvýskytu dvou lexikálních jednotek, jednoduše řečeno, čím častěji se dvě slova vyskytují společně, tím vyšší je jejich MI score, bez ohledu na frekvenci obou jednotek v jazyce, popřípadě ve zkoumaném vzorku jazyka (Oakes 1998, 53; Church – Hanks 1990, 25). Pro výpočet míry asociace dvou slov vyskytujících se ve vzájemných kontextech se používají následující vzorce<sup>16</sup>. První z nich uvádí vzorec pro výpočet hodnoty vzájemné informace (MI score).  $N$  je celkový počet slov v korpusu,  $f(x)$  a  $f(y)$  jsou frekvence členů kolokace v korpusu.

$$mi(x, y) = \log_2 \frac{N \cdot f(x, y)}{f(x) \cdot f(y)}$$

T-score naopak bere v potaz nejprve frekvenci výskytu jednotlivých lexikálních jednotek a až poté frekvenci jejich vzájemného výskytu (Landau 2001, 309). Pro výpočet hodnoty míry kontrastu (t-score) se používá následující vzorec.  $N$  je opět počet slov v korpusu,  $f(x)$  a  $f(y)$  frekvence členů kolokace v korpusu.

$$T = \frac{\left( f(x, y) \cdot \frac{f(x) \cdot f(y)}{N} \right)}{\sqrt{f(x, y)}}$$

---

<sup>16</sup> Zdroj: <http://ucnk.ff.cuni.cz> (duben 2008).



MI score nám může pomoci najít a zkoumat spojení lexikálních jednotek, jejichž souvýskyt, bez ohledu na frekvenci jednotlivých členů spojení, je směrodatný pro fungování daných lexikálních jednotek v jazyce, jde tedy o typická spojení, sousloví, idiomatická spojení. T-score podává informace o lexikálních jednotkách s vysokou frekvencí výskytu, vysokou míru této statistické informace proto vykazují především slova funkční – předložky, spojky, pomocná slovesa apod. T-score je proto vhodnou pomůckou např. pro vyhledání běžných slovních spojení, u nichž frekvence členů spojení je relativně vysoká. MI score napomůže naopak vyhledat spojení řidší, ale typičtější pro členy slovního spojení. Existuje samozřejmě i řada dalších statistických přístupů ke zkoumání vzájemné informace slov vyskytujících se v blízkém sousedství, kombinace výsledků dosažených využitím MI score a t-score jsou ale zřejmě nejjednodušším a nejpraktičtějším způsobem, jak získat základní informace o vzájemném souvýskytu slov v jazyce.

MI score a t-score jsou také základní statistické funkce obsažené v korpusových vyhledávacích většině současných jazykových korpusů. Oba námi sledované korpusy obsahují funkce výpočtu těchto dvou statistických informací a na jejich základě jsou schopny sestavit kolokační tabulky seřazené v závislosti na předvoleném typu požadovaných údajů o míře asociace. V naší analýze Kielipankki a ČNK jsme sledovali kolokační tabulky sestavené na základě obou typů statistických informací a srovnáváním výsledků získaných z obou tabulek jsme v závěrečné fázi získali skupiny kolokačně významných a zajímavých slovních spojení, která by podle nás neměla chybět ve slovníkových heslech zkoumaných substantiv. Kolokační tabulky jsou ale opět pouze výchozím materiálem pro zkoumání jazyka, nelze je chápat jako hotové a univerzálně použitelné výstupy. Už na první pohled je totiž zřejmé, že mnohé kolokace jsou jen konkrétními realizacemi určitých jazykových vzorců (v našem případě např.

spojení číslovky a substantiva času 2 *päivää* – 2dny, 3 *päivää* – 3dny atd.), a je proto nutné je dál třídit.

Kolokační tabulky substantiv času ve finštině (viz přílohy) obsahují pro každé slovo tabulky pro obě výše popsané statistické informace. Vzhledem k obsáhlosti materiálu obsaženého v kolokačních tabulkách vygenerovaných z obou korpusů uvádíme v příloze č. 1 pro každé slovo jen zhruba prvních 50 kolokačních řádků pro každou statistiku, počet řádků u jednotlivých substantiv se liší v závislosti na jednotlivých statistických hodnotách a relevanci konkrétních kolokací (snažili jsme se například o to, aby v tabulkách nechyběly příklady kolokací jednotlivých substantiv se slovesy, v případě kolokací s číslovkami jsme pro jejich velký počet přistoupili k omezování, pokud kolokace vykazovaly nižší hodnoty míry asociace).

Kolokační tabulky finských substantiv času jsme z Kielipankki získali s využitím následujících vyhledávacích parametrů:

1. Kolokace jsme vygenerovali pro jednotlivé tvary – s předvolením možnosti *word*. Jedině tak je možné u jazyka s bohatou tvarovou variací zjistit spojení konkrétních tvarů a tedy kolokace vyskytující se v parole.
2. Kontext pro vyhledávání kolokací jsme omezili na rozmezí  $\pm 3$  slov (wordů) od centrálního slova (KWIC). Zkušebními omezováními kontextu v rozmezí 1-5 slov od KWIC jsme zjistili, že většina kolokací se ve finštině realizuje v bezprostřední blízkosti od KWIC, tedy v kontextu rozmezí  $\pm 1$  slova od KWIC. Přesto jsme pro sestavování kolokačních tabulek použili kontext širší. U substantiva *päivä* jsme z důvodu funkční chyby v konkordančním programu Lemmie, která znemožňuje vyhledávání kolokací

v kontextu napravo od KWIC, vygenerovali kolokační tabulky pouze pro levý kontext.

3. Minimální frekvenci jednotlivých členů kolokačních spojení jsme nastavili na hodnotu 5, minimální hodnotu statistické informace (MI score, t-score) jsme nastavili na hodnotu 5.

## **4.2 Funkční skupiny substantiv vyjadřujících čas ve finštině**

### **4.2.1 Všeobecná substantiva**

O tom, že slova mají ve velké většině více než jeden význam, o lexémové polysémii, toho bylo napsáno již mnoho. Jen málokteré slovo má jen jeden význam a tato skutečnost se odráží i ve slovníkových heslech, v nichž jsou jednotlivé významy většinou hierarchicky uspořádány. O tom, že jednotlivé významy lexémů se realizují většinou až v zapojení do kontextu, ve spojení s jinými slovy, jsme i v této práci mluvili již několikrát. Existují ale i další možné pohledy na fungování jednotlivých významů lexémů v jazyce.

V souvislosti s vysokofrekvenčními substantivy se např. Halliday a Hasan (1976) zmiňují o tzv. „general nouns“ (vše)obecných substantivech, která v jejich pojetí stojí na pomezí lexika a gramatiky. Na jedné straně splňují tato slova všechny podmínky nutné k zařazení do skupiny plnovýznamových, autosémantických slov, na druhé straně plní funkce náležející především slovům gramatickým, vstupují do kohezních a textotvorných vztahů. Hallidayovo a Hasanovo pojetí rozvedla ve svých korpusových analýzách angličtiny např. Michaela Mahlberg (2005).

Mahlberg vymezila několik skupin všeobecných substantiv a zkoumala je na korpusových materiálech. Jednou ze skupin všeobecných substantiv byla pro ni substantiva času. Na základě svých zkoumání pak vytvořila funkčně-významové skupiny jednotlivých okruhů všeobecných substantiv. Teorie obsažená ve dvou výše zmiňovaných pracích byla vypracována na anglickojazyčném materiále. Angličtina je nejen po stránce typologické, ale i po stránce gramatické zcela jiný jazyk než finština, přesto jsme se nechali inspirovat k vytvoření podobného schématu i pro finská substantiva času.

Na vyčlenění jednotlivých funkčně-významových skupin substantiv času mělo vliv několik faktorů. Základními faktory byly faktory významové. Jednotlivé skupiny reprezentují významy užití substantiv vyjadřujících čas, a jejich funkce v různých kontextech, přičemž ne všechny skupiny obsahují všechna námi zkoumaná substantiva. Dalšími důležitými určujícími faktory byly kolokace a spojení, do nichž dané lexémy vstupují, a s tím související jejich tvarová distribuce.

#### 4.2.2 Skupina jednotek času

První funkčně-významovou skupinou je užití časových substantiv jako jednotek času. Ve spojení s číslovkou nebo jiným kvantitativním výrazem vyjadřují některá substantiva času dobu trvání, určitý časový úsek, nebo jsou součástí dalších jednotek označujících fyzikální veličiny. Do této skupiny patří z námi zkoumaných slov všechna substantiva vyjadřující určitou a konkrétní časovou informaci, tedy lexémy *vuosi*, *kuu(kausi)*, *viikko*, *päivä*, *tunti*, *minuutti*, *sekunti*. Tuto skupinu bychom mohli dále rozdělit na přesná numerická vyjádření času (*kaksi vuotta* – dva roky, *kolme minuuttia* – tři minuty, *viisitoista sekuntia* – patnáct sekund/vteřin), přibližné vyjádření času (*pari*

*viikkoa* – pár týdnů, *muutama tunti* – několik hodin) a neurčité vyjádření časového trvání (*monta vuotta* – mnoho let, *useampi viikko* – několik týdnů, v tomto případě by do skupiny patřilo i substantivum *aika* (*paljon aikaa* – hodně času)). Dále sem můžeme zařadit užití časových substantiv jako jednotek fyzikálních veličin času a rychlosti (*5 minuuttia ja 13 sekuntia* – 5 minut a 13 sekund, *kaksi kilometriä tunnissa/metriä sekunnissa* – dva kilometry za hodinu/metry za sekundu).

### 4.2.3 Postpoziční spojení

Další skupinou je spojení substantiva času s postpozicí. Tato skupina je vymezena především po stránce formální, postpoziční spojení vyjadřují různé druhy významů. Protože se ale jedná o obsáhlou skupinu, začlenili jsme ji do našeho pojetí jako skupinu samostatnou. Spojení substantiv času s postpozicí má u všech námi zkoumaných substantiv (a u substantiv ve finštině obecně) statisticky významné hodnoty t-score. Je to dáno hlavně tím, že postpozice<sup>17</sup> jsou ve finštině hojným slovním druhem a postpozičních spojení se užívá pro vyjádření mnoha různých významů, například takových, které odpovídají spojení s předložkami v češtině. Tato funkčně-významová skupina je také jednou z hlavních příčin vysoké frekvence genitivních tvarů u substantiv času (a substantiv vůbec), protože do spojení s většinou postpozic vstupují substantiva (a adjektiva) v tvaru genitivu. Do této skupiny lze zařadit opět všechna časová substantiva vyjadřující konkrétní časové úseky, substantivum *aika* bývá

---

<sup>17</sup> Postpozice je považována, jak uvádí například Lindroos (1997, 13), za slovo gramatické, které ztratilo svou sémantickou nezávislost, a změnilo se v element, který vyjadřuje vztah. Tento vztah vyjadřuje spojení postpozice se substantivem, ale i jeho závislost na slovesu. Finské postpozice („záložky“, proti prepozice – předložky) odpovídají mnoha svými formálními, syntaktickými i sémantickými vlastnostmi českým předložkám, ve spojení s dalšími slovními druhy, především jmény, vyjadřují časové, místní a další vztahy. Na rozdíl od předložek je jejich pozice za slovem, ke kterému se vztahují (odtud prefix post-). I ve finštině se ovšem vyskytují předložky, některé postpozice pak mohou stát jak v pozici za řídícím slovem, tak před ním (např. *ennen* – před).

doplněno o upřesňující kvantitativní výraz (*vähän ajan päästä* – za krátký čas). Významově se tato skupina dá opět dále dělit na postpoziční spojení vyjadřující časové úseky (např. postpozice *aikana, ajan, verran* – *vuoden aikana* – během roku, *kahden vuoden ajan* – v období dvou let, *viikon verran* – týden (po dobu jednoho týdne)), body na časové ose v budoucnosti i minulosti (např. postpozice *sitten, jälkeen, päästä, kuluttua* – *vuoden päästä* – za rok, *päivän jälkeen* – po (jednom) dni, *kuukausi sitten* – před měsícem, *viikon kuluttua* – za týden, *vähän ajan päästä* – za krátký čas, *jonkun ajan jälkeen* – za nějaký čas), pravidelné a opakující se časové úseky (např. *välein* – *minuutin välein* – po minutě, v minutových intervalech). Samostatnou skupinou ve spojení se substantivy času jsou postpozice vzniklé ze substantiva *aika*, o nich jsme se zmínili již dříve.

K předchozí skupině bychom mohli přiřadit i další skupinu vyjadřující časové body na časové ose, tentokrát se ale nejedná o primárně postpoziční spojení, ale o spojení s komparativy časových adverbii, především *myöhemmin* – později a *aikaisemmin* – dříve. Tato spojení jsou v mnoha případech synonymní se spojeními postpozičními a mohou se někdy vzájemně nahrazovat (*viikko myöhemmin* – o týden později x *viikon jälkeen* – po týdnu, *vuosi aikaisemmin* – o rok dříve x *vuosi sitten* – před rokem). Rozdílem v možnosti jejich užití je ale často vztah k okamžiku promluvy. Pokud například mluvíme o události, která se stala v minulosti, a chceme k ní vztáhnout další událost, která se stala rok před ní, užijeme výrazu *vuosi aikaisemmin*, užití spojení *vuosi sitten* vztahuje událost, na kterou odkazujeme, k okamžiku promluvy.

#### 4.2.4 Spojení s adjektivem typu *mittainen*

Další skupinou výrazů, vyjadřujících časové úseky a dobu trvání, jsou spojení substantiv času s adjektivy typu *mittainen*, *pituinen*. Tato adjektiva, odvozená od substantiv nebo jakostních adjektiv, se nevyskytují samostatně, vždy musí být doplněna o upřesňovací kvantitativní výraz v genitivu (*tunnin mittainen esitys*, *muutaman viikon pituinen harjoitusaika*). Takováto spojení jsou alternativou pro české spojení „dvě hodiny dlouhé představení“, „několik týdnů trvající zkušební doba“. Spojení typu *\*kaksi tuntia pitkä esitys*, *\*muutamaa viikkoa pitkä harjoitusaika* nejsou podle výsledků získaných z korpusů pro finštinu typická, v korpusu jsme pro zkoumaná substantiva takový příklad nenašli ani jeden (i v češtině se ale v tomto případě užívá adjektiv *dvouhodinový*, *několikatýdenní*).

#### 4.2.5 Deiktické výrazy

Velkou a v souvislosti s výše zmíněnou teorií všeobecných substantiv relevantní funkční skupinou je skupina výrazů časově deiktických. Časové výrazy mají už ve své podstatě deiktickou povahu, odkazují v čase. O několika možných způsobech časového odkazování jsme se zmínili už v souvislosti s předchozími funkčně-významovými skupinami. Do této skupiny ovšem řadíme taková slovní spojení, u nichž deixe, a to především aktuální deixe, hraje hlavní roli. Jde o spojení substantiv času s deiktickými výrazy (zájmeny, adjektivy), jejichž spojením vzniká aktuální odkaz v čase do budoucnosti, minulosti, či současnosti.

Časový výraz může odkazovat aktuálně, vzhledem k okamžiku promluvy, a to do budoucnosti (adjektivum *ensi* - příští, *ensi vuonna* – příští rok, *ensi viikolla* – příští

týden), do minulosti (adjektivum *viime* – minulý, *viime viikolla* – minulý týden, *viime päivinä* – v minulých dnech) nebo se může vztahovat k přítomnosti (zájmeno *tämä* – tento, *tänä vuonna* – tento rok/letos, *tässä kuussa* – tento měsíc). Libovolnou kombinací těchto deiktických výrazů se substantivy času vzniká pole časových odkazů, které je schopno pokrýt různě velké časové rozměry, ale které není tvarově zcela homogenní (viz kapitola 4.1.2 zabývající se pádovou distribucí u substantiv času). Dalšími možnostmi odkazování je například odkazování na začátek, či konec časového úseku (adjektiva *ensimmäinen* – první, *ensimmäisenä kuun päivänä* – první den v měsíci, *viimeinen* – poslední, *viimeisellä minuutilla* – v poslední minutě), na pořadí na časové ose, vztah časové následnosti a časového předcházení (*seuraava* – další, *seuraavalla viikolla* – další týden, *edellinen*, *edeltävä* – předchozí, *edellisenä päivänä* – předchozí den, *edeltävinä vuosina* – v předchozích letech).

Zajímavá a pro další zkoumání otevřená je otázka textotvorné a kohezní funkce časových substantiv a časových výrazů obecně. Při zkoumání rozsáhlejšího homogenního textu lze vysledovat síť vnitro- i vnětextových odkazů v prostoru a čase, které napomáhají výstavbě a významové soudržnosti textu. Tato struktura je odrazem organizace časoprostorového odkazování v jazyce. Substantiva času jako jedna ze skupin všeobecných substantiv s deiktickými schopnostmi se na organizaci prostoru a času v jazyce podílí velkou měrou. Jen malým dodatkem mohou být v návaznosti na výše zmíněné funkčně-významové skupiny výrazy jako *vuonna 2006* – roku 2006, *viikko numero 11* – týden číslo 11, *kuun 15. päivä* – 15 den v měsíci, které časově (více, či méně přesně) zařazují události na časovou osu a napomáhají tak časovému uspořádání objektivní reality v našem chápání světa.



#### 4.2.6 Idiomatická spojení

Další neméně významnou skupinou je poměrně pestrá a různorodá skupina idiomatických spojení, obsahujících substantiva času. Tuto skupinu nemůžeme zřejmě vymezit zcela jednoznačně po významové stránce, obecně ji lze charakterizovat jako skupinu ustálených časových výrazů. Do této skupiny by pak patřily výrazy s různou mírou idiomatickosti. Lze sem zařadit běžná spojení substantiv času s dalšími slovními druhy (slovesy – *aika kuluu* – čas běží, adjektivy – *pyöreät vuodet* – „kulaté narozeniny, kulatiny“ (*täyttää pyöreitä vuosia*), *sekunnin sadasosa* – setina vteřiny), která stojí na hranici ustálených slovních spojení a idiomatických spojení, i pravé idiomy (podrobně se problematikou časových výrazů ve finštině zabývala ve své diplomové práci například Anna Salminen (2006)). Zkoumání idiomatických slovních spojení substantiv času by bylo zajímavým tématem pro další korpusové analýzy.

Výše vymezené funkčně-významové skupiny v žádném případě nechápeme jako jediné možné a určující rozdělení významů a užití substantiv času v jazyce. Námi načrtnuté skupiny lze dále dělit a organizovat, přičemž zapojením dalších druhů časových výrazů (například časových adverbii) by se toto rozdělení opět poněkud proměnilo. Vymezením funkčně-významových skupin užití substantiv času jsme chtěli ukázat, že jazykové výrazy lze dělit i jinak než pouze formálně či významově a že sledování zapojení jazykových jednotek v parole ukazuje, že propojení a vzájemné působení různých jazykových významů (gramatických, sémantických, kolokačních) je určující pro fungování lexikálních jednotek v jazyce.

## 5 Zpracování materiálů pro slovníková hesla

### 5.1 Zpracování materiálů

Jak jsme již zmínili v úvodní části práce, jedním z cílů této studie je využít korpusové materiály získané z Kielipankki a ČNK k sestavení návrhů slovníkových hesel pro nový dvoujazyčný finsko-český slovník. Námi vymezená skupina slov není nijak velká, korpusový materiál, který jsme ke svým zkoumáním využili, byl ovšem více než obsáhlý. Z hojného statisticky i ručně protříděného materiálu jsme získali hodnotné a obsáhlé výstupy využitelné pro sestavení slovníkových hesel.

Při lexikografickém zpracování korpusových materiálů jsme v první řadě přihlíželi k povaze slovníku, ve kterém by mohla námi navržená hesla být obsažena. Dvoujazyčný překladový slovník dvou malých jazyků by měl mít rozsah malého středního až středního slovníku, tedy zhruba v rozmezí od 20 000 do 30 000 hesel pro každou část. Pokud se jedná o tištěný slovník, je třeba vzhledem k omezenému rozsahu využívat prostor ekonomicky a používat systém zkratk a odkazů, aby nedocházelo ke zbytečnému opakování stejných informací. Výhodné je také tzv. hnízdování – přidružování slov odvozených, složených či jinak příbuzných k hyperlemmatu, což kromě úspory místa také zjednodušuje vyhledávání. Konkrétní podoba slovníkových hesel by pak záležela na skutečném rozsahu, formátu a grafickém zpracování slovníku.

Při sestavování slovníkových hesel jsme vycházeli od zdrojového jazyka a k vybraným výrazům jsme dohledávali nejvhodnější ekvivalenty v jazyce cílovém. Jak už bylo řečeno výše, základním zdrojem pro výběr a sestavování slovníkového hesla byly kolokační tabulky jednotlivých slov, frekvence jednotlivých jejich realizací (tvarů, popřípadě ve finštině také důležitých kompozit) a v neposlední řadě konfrontace statistických korpusových informací s existujícími slovníky a naším povědomím o obou

jazycích. Ani práce s korpusovými materiály a lexikografickými příručkami se, jak uvádí např. Sinclair (1991, 38), neobejde bez ručního třídění a zpracování na základě intuice rodilého mluvčího, či cvičeného uživatele jazyka. Statistické informace sice mohou ukázat, co je v jazyce časté a typické, pouhá kvantita ale neznamená vždy relevantnost a důležitost. Sinclair (1991, 69) poukazuje také na to, že vedle čistě statistických informací o výskytu slov je třeba brát na vědomí také to, že v jazyce existují gramatická pravidla, která mohou počet výskytů určitého jevu silně ovlivnit a statistické informace tak zkreslit. Takovým příkladem ve finštině může být například už několikrát zmiňovaný případ postpozic, které od slova, jež určují, vyžadují jistý gramatický tvar. Z tohoto důvodu pak substantiva času, stejně jako ostatní skupiny substantiv ve finštině vykazují vysoké procento výskytu genitivních tvarů. Funkce genitivu je přitom v tomto případě, ale i v mnoha jeho dalších užitích, hlavně gramatická.

Námi navržená slovníková hesla mají sloužit především jako model možného zpracování a možné úpravy slovníkových hesel. Například část materiálů, obsahující příklady užití, by mohla být rozsáhlejší a mohla by obsahovat příklady užití v širším kontextu. Vzhledem k omezenému prostoru v tištěném slovníku je podle našeho názoru užívání minimálního kontextu vhodnější. Další možností by bylo ještě větší využití hnízdování slov, zařazení i dalších odvozených slov, či odkazy na další hesla. Všechny tyto případy by ovšem závisely na konkrétním konceptu slovníku, na jeho formátu a grafickém zpracování. Informace o frekvenci výskytu jednotlivých substantiv v jazyce jsou důležité pro určení relevance daného slova pro uživatele. Pokud má slovo vysokou frekvenci výskytu, znamená to, že je často užíváno a je tedy svým způsobem pro uživatele jazyka důležité. Pokud odhlédneme od obecné otázky vztahu frekvence slov v jazyce a jejich gramatických funkcí a významů (nejfrekvencovanější slova jsou

většinou slova gramatická, v češtině pak například předložky, pomocná slovesa, ve finštině postpozice apod.) a soustředíme se na skupinu substantiv času, můžeme si na základě údajů o frekvencích výskytu substantiv času ve finštině (v Kielipankki a Frekvenčním slovníku finštiny) uvedených v tabulce č. 4 povšimnout rozdílů mezi jejich hodnotami. Vysokou frekvenci vykazuje v obou zdrojích substantivum *vuosi*, pokud bychom se soustředili na situaci v korpusu, vzájemně srovnatelně vysoké frekvence vykazují substantiva *aika*, *päivä* a *viikko*. Naopak například substantiva *kuu*, *minuutti* a *sekunti* mají v porovnání s ostatními poměrně nízkou frekvenci výskytu (znovu podotýkáme, že pouze v porovnání s ostatními substantivy času, jinak i tato substantiva patří mezi frekvenčně významná slova).

Vyšší frekvence výskytu ale vždy nemusí znamenat, že pro dané slovo má být ve slovníku vyhrazeno odpovídajícím způsobem více prostoru. Některá slova jsou frekventovaná, ale jejich význam je omezen na několik málo užití a kolokačních spojení. Z vysoké frekvence substantiva *vuosi* například velkou část výskytů zaujímají určení konkrétních roků/letopočtů (*vuonna 2000*). Jiná méně frekventovaná slova mají naopak celou řadu různých významů a realizací (např. substantiva *kuu* a *kuukausi*). Dalším hlediskem, ke kterému jsme přihlíželi při vytváření návrhů slovníkových hesel, byly tedy konkrétní skupiny výskytů zkoumaných lexémů v jazyce, jejich zapojení do kolokací a širě realizací jejich významů. Otázku významů a funkčně-sémantických vztahů, do nichž substantiva času vstupují, jsme podrobněji popsali v kapitole 4.2.

## 5.2 Finsko-český slovník

Ve finsko-české části slovníková hesla obsahují základní gramatické informace (kromě základního tvaru nominativu singuláru jsou v závorce uvedené tvary genitivu

singuláru, partitivu singuláru a partitivu plurálu)<sup>18</sup>. Následují základní ekvivalenty v cílovém jazyce. Na dalších řádcích jsou pak základní a podle korpusu nejhojnější výskyty (zařazené do minimálního kontextu) následované kolokacemi se slovesy, dalšími frekventovanými spojeními a výskyty (u substantiva *aika* postpozicemi odvozenými od substantiva *aika*). Následují další slova odvozená od základního substantiva (adverbia, adjektiva). Další částí jsou typická spojení a idiomy, poslední částí jsou pak kompozita utvořená od základového substantiva. Podoba a rozsah hesel jednotlivých substantiv se liší v závislosti na jejich relevanci pro slovník (čím frekventovanější slovo a čím větší významová pestrost, tím obsáhlejší heslo a tím kompletněji také realizována výše popsaná struktura hesla).

Významnou skupinou výskytů časových substantiv v jazyce jsou kompozita. Tvoření kompozit je pro finštinu typické a skládání slov je jedním z hlavních zdrojů obohacování slovní zásoby ve finštině. Proto i skupina kompozit substantiv času je rozsáhlá. Při sestavování slovníkových hesel nastává v souvislosti s kompozity otázka, zda a jaká kompozita mají být obsažena ve slovníkovém hesle, popřípadě v heslovém hnízdě jednotlivých lexémů.

Z finského korpusu jsme vygenerovali kompozita substantiv času tak, že jsme vyhledávali slova (kompozita), jejichž složkou jsou námi zkoumaná substantiva času a jejich různé tvary. Získali jsme tak dílčí seznamy kompozit obsahujících substantiva času, tyto dílčí seznamy jsme pak dále třídili. Z formálního hlediska existují tři možnosti umístění substantiva času jako částí kompozita, a to v pozici na začátku, uprostřed (v případech, kdy kompozitum obsahuje více než dvě složky) nebo na konci

---

<sup>18</sup> Nominativ singuláru jako obecně přijímaný základní tvar slova je reprezentantem celého lexému. Další tvary jsou kromě své vysoké frekvence ve finštině vybrány hlavně jako pomůcky pro tvorbu dalších tvarů slova. Od genitivu singuláru lze odvodit tzv. slabý samohláskový kmen nutný pro tvorbu většiny pádů ve finštině. Partitiv singuláru ukazuje naopak tzv. silný samohláskový kmen, popřípadě souhláskový kmen substantiva, partitiv plurálu je zase výchozí tvar pro tvorbu tvarů množného čísla.

složeného slova. U slovníkových hesel je v případě kompozit problémem také to, zda mají být do slovníkového hnízda konkrétního lexému umístěna všechna relevantní kompozita, či zda se do slovníkového hesla mají umístit pouze ta kompozita, v nichž je substantivum času v pozici na začátku slova. Ostatní případy by se řadily abecedně k odpovídajícím lemmatům, podle počáteční složky kompozita. V případě substantiva *aika* by pak součástí hesla byla slova jako *aikataulu*, či *aikakausi*, naopak slova jako *työaika* nebo *vuodenaika* by se ve slovníku řadila k heslům *työ*, resp. *vuosi*. Takového rozdělení respektuje formální stránku slov a pro uživatele zpřehledňuje vyhledávání slov podle abecedního pořádku ve slovníku, proto k němu v návaznosti na podobné uspořádání ve starším *Suomi-tšekki-suomi taskusanakirja* (Lindroos 1984) přistupujeme i v této práci.

V samotných návrzích slovníkových hesel tedy uvádíme jen ta kompozita, která mají časové substantivum na pozici první ze svých složek, ostatní kompozita přesto uvádíme v příloze jako materiálové doplnění informací o zkoumaných substantivech. Substantivum *aika* vstupuje do kompozit v několika tvarových modifikacích. Kromě nominativního tvaru singuláru *aika* (*aikakausi*) se v kompozitech vyskytují i další tvary, a to především genitiv singuláru *ajan* (*ajankohta*). Výjimečným příkladem užití illativu singuláru substantiva *aika* je složené sloveso *aikaansaada*.

### 5.2.1 Substantivum *aika*

Substantivum *aika* má ve finštině dva základní významy. Finština rozeznává jednak příslovce míry *aika* ve významu „celkem, dost“, jednak námi zkoumané substantivum *aika* ve význam čas, období. Obsahem našeho zájmu bylo pouze substantivum času. Ze všech zkoumaných časových substantivum je substantivum *aika*

významově nejabstraktnější, jeho význam neurčuje na rozdíl od ostatních substantiv této skupiny žádnou konkrétní časovou hodnotu.

Škála realizací a kolokačních zapojení substantiva *akia* je celkem pestrá. V případě kolokací vypočtených s použitím statistické informace t-score se na horních místech v kolokační tabulce vyskytují kolokace tvarů *ajan*, *aikana*, *aikaan*. Těmito gramatizací tvarů substantiva *aika* vzniklými postpozicemi jsme se zabývali už v předchozích kapitolách. Ačkoli příbuznost se substantivem *aika* je stále velmi zřetelná, a ačkoli v samotném korpusu jsou tyto postpozice otaggovány jako tvary substantiva *aika*, v současných finských gramatikách jsou považovány za samostatná slova. Otázkou tedy je, zda mají i v překladovém slovníku stát samostatně, či zda mají být včleněny do hesla pro substantivum *aika* jako jeho specifické tvary. Přestože se gramaticky a zčásti i sémanticky (jde slova z velké části gramatizovaná) vyčleňují jako samostatné výrazy, jejich význam je stále silně spojen s časovým určováním, a proto je zřejmě vhodné je zařadit k lexému *aika*. Navíc ani *Suomi-tšekki-suomi sanakirja*, ani výkladový slovník *Suomen kielen perussanakirja* nemají pro tyto postpozice samostatná hesla.

Substantivum *aika* je vedle výše zmíněných pospozic také zdrojem mnoha dalších časových výrazů. Původem tvary substantiva *aika* jsou mnohá časová adverbia, například adverbia *ajoissa* – včas (*tulla ajoissa* – přijít včas), *aika ajoin* – občas jsou lexikalizované tvary substantiva *aika*, které dnes plní funkci nezávislých lexémů. Jejich formální, ale stále také významová souvislost se substantivem *aika* je ale stejně jako u postpozic tak silná, že i tato adverbia jsou často, a to i v našem případě, kladena ve slovnících do hesla substantiva *aika*.

Substantivum *aika* vstupuje velmi často do různých poloidiomatických a idiomatických spojení. Takováto spojení jsou sice pro členy spojení typická, a pokud se vyskytují v korpusu, mají vysoké hodnoty především MI score, protože jejich frekvence ale není většinou příliš významná, jsou z korpusů, také kvůli statistickým omezením, která se většinou využívají při sestavování kolokačních tabulek jako prostředek k omezení výskytu nechtěných náhodných spojení slov s nízkou frekvencí, ale s vysokou hodnotou vzájemné informace, těžko vysledovatelná. Přesto se i při běžných nastaveních dají v kolokačních tabulkách nalézt některá idiomatická spojení, která jsou svým častějším výskytem zajímavá také pro uživatele slovníku. Příkladem takového idiomatického spojení je výraz *korkea aika* – nejvyšší čas (oproti češtině je zde adjektivum *korkea* – vysoký užito v prvním stupni), nebo spojení *kestää iät ajat* – trvat věčnost (doslova \*trvat věky časy). Podobné výrazy s určitým stupněm idiomatickosti jsou pro slovník vhodným příkladem běžného, ale něčím specifického užití jazyka. Pokud se jejich ekvivalenty v cílovém jazyce liší tak, že je nelze volně přeložit, ani analogicky utvořit, jsou také užitečným příkladem v obou jazycích výrazově podobných, ale formálně odlišných výrazů.

#### **AIKA (ajan, aikaa, aikoja) – čas, doba, období (aikakausi)**

*paljon, vähän aikaa* – hodně, málo času, *koko ajan* – stále, *eri aikoina* – v různou dobu, *ajan mitaan* – časem; *viettää aikaa* – trávit čas, *aika kuluu* – čas běží, *kestää jonkin aikaa* – trvat nějakou dobu; *jhk. aikaan* – v nějaký čas, někdy, *siihen aikaan* – v ten, onen čas, tehdy, *12 aikaan* – ve 12 hodin, *keskiyön aikaan* – kolem půlnoci, *samaan aikaan* – ve stejnou dobu, stejně; *jnak. aikana* – v nějaké době, za časů, během, *sinä aikana* – v té době, *sodan aikana* – během války, *meidän aikanamme* – za našich časů; *ajoiissa* – včas, *aika ajoin* – občas; *korkea aika* – nejvyšší čas, *kestää iät ajat* – trvat věčnost, *kovat ajat* – těžké časy, *muutosten aika* – čas změn, *paikallinen aika* – místní čas, *se vie aikaa* – to chce čas; *vanhoina aikoina* – za starých časů, *kaikkien aikojen ennätys* – historický rekord; *aikaansaada* – způsobit, uskutečnit, *aikakausi* – období, věk, éra, *aikakaulehti* – časopis, *aikataulu* – časový plán, rozvrh

Další kompozita substantiva *aika*:

*elinaika* – doba života, život (časově), *irtisanomisaika* – výpovědní lhůta, *keskiaika* – středověk, *koeaika* – zkušební doba, *kulta-aika* – zlaté časy, zlatý věk, *määräaika* –



lhůta, *nykyaika* – současnost, přítomnost, *odotusaika* – čekací doba, *syntymäaika* – datum narození, *työaika* – pracovní doba, *vapaa-aika* – volný čas, *voimassaoloaika* – doba platnosti, *vuodenaika* - roční období

### 5.2.2 Substantivum vuosi

Substantivum *vuosi* vyjadřuje ve finštině ve spojení s číslovkami časovou jednotku označující jeden rok (12 měsíců) *18 vuotta vanha* – 18 let starý nebo odkazuje na konkrétní roky *vuonna 2005* – roku 2005, ve spojení s deiktickými výrazy a postpozicemi pak slouží k odkazování v čase *ensi vuonna* – příští rok, *vuoden päästä* – za rok. Podle Frekvenčního slovníku finštiny i frekvence v Kielipankki je nejfrekventovanějším substantivem času, což je v souvislosti s tradicí počítání času v naší společnosti způsobeno jeho hojným užíváním v odkazování na konkrétní letopočty.

Přes vysokou frekvenci není sémantické pole substantiva *vuosi* příliš bohaté. Vedle výše zmíněných výskytů a významů se objevuje v idiomatických spojeních jako *täyttää pyöreitä vuosia* – slavit kulaté výročí, narozeniny nebo *vuosi vuodelta*. Typickými kompozity, jejichž význam souvisí opět především s počítáním času, jsou *vuosikymmen* – desetiletí, *vuosisata* – století, *vuosituhä* - tisíciletí, *vuosiluku* - letopočet. V neposlední řadě je třeba zmínit kompozitum *uusivuosi*, které psáno s malým počátečním písmenem a dohromady je ekvivalentem českého Nový rok (první den v roce), oproti *uusi vuosi* – nový rok (*Onnelista uutta vuotta!* – Šťastný nový rok!)

#### VUOSI (vuoden, vuotta, vuosia) - rok

*vuosi sitten* – před rokem, *vuoden päästä/kuluttua* – za rok, *ympäri vuoden* – během roku; *vuonna 2005* – roku 2005, *kerran vuodessa* – jednou za rok, *25 vuotta vanha* – 25 let starý, *pěťadvacetiletý*; *viime vuonna* – vloni, *tänä vuonna* – letos, *ensi vuonna* – příští rok, *joka vuosi*, *vuosittain* – každý rok; *juhlia vuosia* – slavit výročí; *täyttää pyöreitä vuosia* – slavit kulaté výročí, narozeniny, *vuosi vuodelta* – rok co rok; *vuosiluku* - letopočet, *vuosisata* – století, *vuosituhä* – tisíciletí, *Uusivuosi* – Nový rok

Další kompozita substantiva *vuosi*:

*lukuvuosi* – semestr

### 5.2.3 Substantiva *kuu* a *kuukausi*

Mezi substantivy *kuu* a *kuukausi* panuje vztah synonymity, substantivum *kuukausi* je vlastně složeninou slov *kuu* (měsíc) a *kausi* (období) a jeho význam se dá z jeho stavby odvodit jako období (jednoho) měsíce. Synonymita těchto dvou substantiv ovšem není absolutní. Substantivum *kuu* má vedle významu měsíc jako časová jednotka také význam měsíc jako vesmírné těleso, luna. V kapitole 4.1.2 jsme se podrobněji zabývali rozdíly v užívání těchto dvou substantiv. Substantivum *kuu* se, jak vyplývá z korpusových dat, užívá především ve spojení s deiktickými slovy a takováto spojení pak odkazují na konkrétní měsíce (*ensi kuussa*, *viime kuussa*), substantivum *kuukausi* se pak užívá ve významu měsíce jako časového úseku zhruba 30 dní, např. ve spojení s číslovkami (*kolme kuukautta*, *kahden kuukauden päästä*). V některých případech lze ale užít obou substantiv (*kerran kuussa/kuukaudessa*) aniž by se změnil význam spojení. Jsou to především případy, kdy toto časové substantivum vyjadřuje právě časový úsek jednoho měsíce, zdá se, že v těchto případech se užití a význam etymologicky staršího slova *kuu* stále ještě překrývá s významem kompozita *kuukausi*.

V některých případech, kdy lze užít obou substantiv ve stejném kontextu pak může dojít ke sporu mezi názory mluvčích na to, zda dané spojení je pro finštinu přirozené, či nikoli. Například místo výrazu *olla 6. kuulla raskaana* je podle vyjádření, která jsme získali od rodilých mluvčích, přirozenější užívat spojení *olla 6. kuukaudella raskaana*, i když první ze spojení vykazuje v Kielipankki vysokou hodnotu MI score.

Vzhledem ke své časové motivovanosti pak do kompozit s časovým významem vstupuje z těchto dvou substantiv především substantivum *kuukausi*. V takovýchto

spojeních pak první složka složeniny (*kuukausi*-) většinou časově určuje další složky *kuukausipalkka*, *kuukausilippu* ap. Názvy měsíců jsou pak ve finštině složeninami se základovou složkou substantiva *kuu* – *tammikuu*, *helmikuu* atd.

### **KUUKAUSI (kukauden, kuukautta, kuukausia) – měsíc**

*12 kuukautta* – 12 měsíců, *kuukausi sitten* – před měsícem, *kuukauden aikana* – během měsíce, *viime kuukausina* – v posledních měsících; *kuukausittain* – měsíčně; *kuukausiliite* – měsíční příloha, *kuukausilippu* – měsíční jízdenka, *kuukausipalkka* – měsíční plat, mzda, *kuukausitulot* – měsíční příjmy; *kuukautiset* – měsíčky (menstruace)

### **KUU (kuun, kuuta, kuita) – měsíc**

*tässä kuussa* - tento měsíc, *ensi, viime kuussa* - příští, minulý měsíc; *200 euroa kuussa* – 200 eur za měsíc, *kerran kuussa* – jednou měsíčně, *kuun vaihteessa* – na přelomu měsíce; *6. kuulla raskaana* – v šestém měsíci (těhotenství)

Další složeniny substantiva *kuu*:

*tammikuu* – leden, *helmikuu* – únor, *maaliskuu* – březen, *huhtikuu* – duben, *toukokuu* – květen, *kesäkuu* – červen, *heinäkuu* – červenec, *elokuu* – srpen, *syyskuu* – září, *lokakuu* – říjen, *marraskuu* – listopad, *joulukuu* - prosinec

## **5.2.4 Substantivum *viikko***

Substantivum *viikko* patří do skupiny slov přejatých. Označuje úsek jednoho týdne a podobně jako u ostatních časových substantiv se tento význam realizuje v kolokacích s dalšími slovy jako časový úsek *kestää kaksi viikkoa* – trvat dva týdny nebo odkaz na bod na časové ose *kaksi viikkoa sitten* – před dvěma týdny.

Mezi kompozita tohoto substantiva patří také složeniny původně převzaté (především ze švédštiny) a pofinštěné např. *keskiviikko*- středa nebo *viikonloppu* – víkend.

## VIKKO (viikon, viikkoa, viikkoja) – týden

*tällä viikolla* – tento týden, *viikko sitten* – před týdnem, *kaksi viikkoa* – dva týdny, 14 dní, *kerran viikossa* – jednou týdně; *viikottain* – týdně, *viikolla* – v týdnu; *viikonloppu* – víkend

Další kompozita substantiva *viikko*:

*keskiviikko* – středa, *työviikko* – pracovní týden, *alkuviikko* – začátek týdne

### 5.2.5 Substantivum *päivä*

Substantivum *päivä* patří mezi substantivy času k těm frekventovanějším. Jeho českému ekvivalentu den ve finštině odpovídá kromě *päivä* také substantivum *vuorokausi*, které se ve finštině používá k označení dne jako 24 hodin. Výraz *tänä päivänä* označuje podobně jako v češtině dnes (ve fin. také *tänään*) jednak dnešek jako dnešní den, ale také dnešek jako současnost. Z výrazů idiomatických vykazuje v korpusu vysokou hodnotu MI score spojení *yötä päivää* (ve dne v noci).

Substantivum *päivä* má bohaté pole kompozit. Jedná se například o výrazy označující části dne (*aamupäivä* – dopoledne, *keskipäivä* – poledne, *iltapäivä* – odpoledne), svátky a významné dny (*juhlapäivä* – svátek, *merkkipäivä* – významný den, *äitienpäivä* – Den matek), popřípadě o složeniny, v nichž *päivä* jako první část kompozita blíže určuje jeho další části *päivälehti* (proti *iltahelti*), *päiväunet* (proti *yöunet*).

## PÄIVÄ (päivän, päivää, päiviä) – den

*samana päivänä* – ve stejný den, *tänä päivänä* – dnes, v současnosti, *keskellä päivää* – uprostřed dne, *pitkin päivää* – přes den, *jonain päivänä* – jednoho dne, jednou; *avoimien ovien päivä* – den otevřených dveří, *yötä päivää* – ve dne v noci; *päiväkirja* – deník, *päiväkoti* – mateřská škola, *päivälehti* – deník (noviny), *päivämäärä* – datum, *päivällinen* – oběd, *ottaa päiväunet* – dát si šlofika

Další kompozita substantiva *päivä*:

*aamupäivä* - dopoledne, *keskipäivä* - poledne, *iltapäivä* – odpoledne, *arkipäivä* – všední den, *joulupäivä* – 1. svátek vánoční (25. prosince), *juhlapäivä* – svátek, *merkkipäivä* – významný den, *nykypäivä* – dnešek, současnost, *syntymäpäivä* - narozeniny, *työpäivä* – pracovní den, *uudenvuoden päivä* – Nový rok, *ystävänäpäivä* – den sv. Valentýna, *äitienpäivä* – den matek

### 5.2.6 Substantivum *tunti*

Substantivum *tunti* jako výraz pro jednotku času se kromě pro substantiva času typických spojení s číslovkami vyskytuje také jako součást názvu jednotky rychlosti *50 kilometriä tunnissa* – 50 kilometrů za hodinu. *Tunti* jako časový úsek má pak více významů. Prvním významem je hodina jako 60 minut *matka kesti kaksi tuntia* – cesta trvala dvě hodiny, dále pak vyučovací hodina (s rozličnou délkou, ve škole většinou 45 minut) *Meillä on tänään vain 4 (oppi)tuntia* – Máme dnes čtyři hodiny, *matematiikan tunti* – hodina matematiky, *tanssitunti* – hodina tance.

Kompozita substantiva *tunti* pak opět souvisejí s jeho dvěma výše zmíněnými významy. Vedle již zmíněného *tanssitunti* např. *oppitunti*, ve významu hodiny jako časového úseku pak například *tuntikausi*, při vyjadřování času pak *tasatunti* – celá hodina.

#### **TUNTI (tunnin, tuntia, tunteja) – hodina**

*puoli tuntia* – půl hodiny, *tunnin päästä* – za hodinu, *esitys kesti tunnin* – představení trvalo hodinu, *50 kilometriä tunnissa* – 50 kilometrů za hodinu, *24 tuntia vuorokaudessa* – 24 hodin denně; *tuntikausi* - hodina, *tuntipalkka* – hodinová mzda

Další kompozita substantiva *tunti*

*oppitunti* – hodina (výuky), *tasatunti* – celá hodina

### 5.2.7 Substantiva *minuutti* a *sekunti*

Substantiva *minuutti* a *sekunti* jako jednotky přesného vyjadřování času jsou podobně jako substantivum *viikko* slova původem převzatá. Jejich základním významem je ve spojení s číslovkou vyjadřování časového určení nebo časového úseku *kolme minuuttia seitsemän sekuntia* tři minuty a sedm sekund, substantivum *sekunti* je také součástí výrazu jednotky rychlosti *5 metriä sekunnissa* – 5 metrů za sekundu.

Z kompozit je pro obě substantiva společné kompozitum *sekunti-* resp. *minuuttiviisari* – vteřinová resp. minutová ručička. Substantivum *sekunti* se vyskytuje v kolokacích označujících části sekundy *sekunnin kymennys*, *sadasosa* – desetina, setina vteřiny.

#### **MINUUTTI (minuutin, minuuttia, minuutteja) – minuta**

*muutama minuutti* – několik minut, *minuutin välein* – v minutových intervalech; *viimeisellä minuutilla* – v poslední minutě; *minuuttiviisari* – minutová ručička

#### **SEKUNTI (sekunnin, sekuntia, sekunteja) – sekunda, vteřina**

*pari sekuntia* – pár vteřin, *metriä sekunnissa* – metrů za sekundu; *viimeisillä sekunneilla* – v posledních vteřinách, *sekunnin kymennys* – desetina vteřiny, *sekunnin sadasosa* – setina vteřiny; *sekuntiviisari* – vteřinová ručička

## 5.3 Česko-finský slovník

Návrhy slovníkových hesel v česko-finské části opět vycházejí od zdrojového jazyka, materiály z korpusu SYN2000 byly doplněny o vhodné odpovídající výrazy ve finštině z Kielipankki, materiál byl poté upraven konfrontací se Suomi-tšekki-suomi taskusanakirja a Suomen kielen perussanakirja. Podoba a formát slovníkových hesel opět nejsou definitivní a záležely by na formátu celého slovníku. Nejprve je uvedeno lemma ve výchozím jazyce doplněné o tvar genitivu singuláru a gramatický rod.

Následují základní finské ekvivalenty. Dále uvádíme frekvenčně významné příklady užití, spojení se slovesy a výběrové příklady idiomatických spojení. Vzhledem k tomu, že čeština neužívá kompozita tak často, jako finština, v česko-finských heslech chybí informace o slovech složených. Mohla by být ale doplněna například slova odvozená (adjektiva, deminutiva apod.).

Základními materiály při tvorbě nábrhů slovníkových hesel pro česko-finskou část pro nás byly korpusy SYN2000 s využitím korpusového manažeru Bonito i v online verzi Word Sketch Engine (WSE). Vedle kolokačních tabulek vygenerovaných v programu Bonito jsme využívali funkce WSE Word Sketch a Thesaurus.

Jak uvádí např. Čermák (Blatná – Čermák 1995, 231) cizí jazyk, do kterého se český jazyk slovníkově převádí, by měl pokrývat rodilý mluvčí tj. cizinec. Nerodilý mluvčí nemůže ani nejlepším studiem nikdy dosáhnout dokonalosti jazykového citu rodilého mluvčího, a proto by v případě zpracování jazyka do slovníkových příruček ten, kdo zpracovává cílový jazyk pro slovník, měl být rodilým mluvčím daného jazyka. Přesto i jako nerodilý mluvčí finštiny přistupujeme v této práci ke zpracování obou částí slovníku, jsme si vědomi, že v části česko-finské se musíme spoléhat pouze na své omezené znalosti cílového jazyka.

Protože v obou částech slovníku vycházíme od zdrojového jazyka, části si přímo neodpovídají, odrážejí pouze situaci zdrojového jazyka v korpusu a v některých fázích jsou pouze doplněny materiály z existujících slovníků tak, aby podávaly informace o daných lexémech v pokud možno co nejširším rozsahu. Tam kde podle nás informace není relevantní pro dané slovníkové heslo, ale je vhodná k doplnění popisu daného lexému, připojujeme poznámky.

Kombinací funkcí korpusových manažerů Bonito a WSE se dají získat pestřejší informace o zkoumaných slovech, než pouze z kolokačních tabulek a jednoduchých konkordancí v případě Kielipankki, česko-finská část proto obsahuje více informací například o typických vazbách substantiv času se slovesy, nebo také větší počet příkladů idiomatičtějších spojení substantiv času. Zdánlivě větší obsahová bohatost hesel v česko-finské části je ve finsko-české části vyvážena větším důrazem na rozdíl ve významech užití jednotlivých tvarů zkoumaných substantiv času, systém tvarů finských substantiv je totiž v porovnání s českým pádovým systémem u substantiv sémanticky jinak uspořádan, významům některých pádů ve finštině odpovídají v češtině například předložková spojení.

### 5.3.1 Substantivum čas

Kombinací výsledků výpočtu kolokačních tabulek v korpusovém manažeru Bonito a funkce Word Sketch v online manažeru Word Sketch Engine jsme získali vyčerpávající informace o kolokacích zkoumaných substantiv času jak s dalšími vysokofrekvenčními především gramatickými slovy (předložkami, spojkami) s vysokou hodnotou t-score, tak typická spojení s adjektivy, slovesy a dalšími substantivy. Spojení se spojkami můžeme považovat spíše za spojení ovlivněná postavením daných slov ve větě, a tedy náhodnými pozicemi ve vzájemné blízkosti. Spojení s předložkami už mají i sémantickou motivaci, o sémantice předložek a předložkových spojeních v češtině v této práci ovšem podrobněji nepojednáváme.

Zajímavou a pro uživatele slovníku důležitou skupinou kolokací jsou typická spojení substantiv času s adjektivy, slovesy a dalšími substantivy. Mezi typické kolokace substantiva *čas* s adjektivy patří spojení *volný čas*, se sportovní tematikou související spojení *nastavený čas*, či *oddechový čas*, dále např. *ztracený čas*, *letní a*



*zimní čas* nebo *pravý čas* (v *pravý čas*). Mezi slovesa vyskytující se v kolokacích se substantivem *čas* patří především slovesa *plynout* a *letět* (v souvislosti s časem jako pohybující se entitou – *čas plyne, čas letí*), v souvislosti s využíváním času pak slovesa *mít, trávit, šetřit, či marnit, čas se dá samozřejmě také měřit*.

Ze sémantických polí substantiv získatelných za pomoci funkce Word Sketch lze pak poměrně snadno získat také některá typická idiomatická spojení, například z kolokace substantiv *čas* a *peníze* lze usuzovat, že daná kolokace je realizována především jako spojení *čas jsou peníze*, podobně kolokace adjektiv *starý* a *dobry* ukazuje ke spojení *staré dobré časy*. Dalšími typickými idiomatickými spojeními pak mohou být např. *nejvyšší čas*, spojení *na věčné časy* apod.

#### ČAS(u) m – aika, aikamuoto

*před časem* – jonkun ajan sitten, *jednou za čas* – aika ajoin, *časem* – ajan mitaan; *trávit* č. – viettää aikaa, *marnit* č. – tuhata aikaa, *měřit* č. – mitata aikaa, č. *plyne* – aika kuluu, č. *letí* – aika rientää; *volný* č. – vapaa-aika, *letní/zimní* č. – kesäaika/talviaika; *čas jsou peníze* – aika on rahaa, *pracovat přesčas* – tehdä ylittöitä, *staré dobré časy* – vanhat hyvät ajat (za *starých dobrých časů* – vanhoina hyvinä aikoina), *za našich časů* – meidän aikanamme, *nejvyšší čas* – korkea/korkein aika, *na věčné časy* – ikuisiksi ajoiksi

### 5.3.2 Substantivum rok

Substantiva vyjadřující konkrétní časové údaje se ve spojení s předložkami, podobně jako spojení finských substantiv s postpozicemi, užívají k vyjadřování časových údajů, úseků, či odkazování v čase (*před rokem, za rok*). Z adjektiv se substantivum *rok* pojí například se slovy *loňský, minulý, letošní, či příští*. Typickými kolokacemi jsou pak např. *přestupný rok, světelný rok, školní rok*. Z lexikografického hlediska je poněkud problematická otázka množného čísla substantiva *rok*. Vedle plurálu *roky* se totiž užívá (hlavně při odkazování na minulost) také plurálního výrazu *léta* (*před dvěma lety*). Problémem pak je, zda se tyto výrazy mají řadit k lemmatu *rok*,

či k lemmatu *léto* (s nímž souvisejí formálně). My příklad takového užití uvádíme jako součást lemmatu *rok* s poznámkou, že se jedná nesystémové užití plurálu slova *léto*.

#### **ROK(u) m – vuosi**

*za rok* – vuoden päästä, kuluttua, *před rokem* – vuosi sitten; *loňský, minulý rok* – viime vuosi/vuonna, *příští rok* – ensi/seuraava(na) vuosi/vuonna, *letošní rok* – tämä vuosi, *letos* – tänä vuonna, *každý rok* – joka vuosi; *roku 2000* – vuonna 2000, *je o dva roky starší* – hän on kaksi vuotta vanhempi; *přelom roku* – vuoden vaihto, vaihde, *přestupný rok* – karkausvuosi, *světelný rok* – valovuosi, *školní rok* – kouluvuosi, *Nový rok* – uusivuosi, *rok od roku* – vuosi vuodelta, *jednou za uherský rok* – harvoin

*devadesátá léta 20. století* – 1990-luku

### **5.3.3 Substantivum *měsíc***

Podobně jako ve finštině má substantivum *měsíc* i v češtině více významů, čeština má ale na rozdíl od finštiny pro všechny tyto významy pouze jedno substantivum. Vedle významu časové jednotky (zhruba 30 dní) jde pak o významy části roku (*měsíc leden*) a samozřejmě také o název vesmírného tělesa (s velkým počátečním písmenem jde o *Měsíc* jako přirozenou družici země, s malým počátečním písmenem pak *měsíc* jako typ vesmírného tělesa).

#### **MĚSÍC(e) m – kuukausi, kuu**

*před měsícem* - kuukausi sitten, *během měsíce* - kuukauden aikana, *za měsíc* – kuukauden päästä, kuluttua; *kalendářní m.* – kalenterikuukausi, *měsíční výdělek* – kuukausipalkka, *být v pátém měsíci (těhotenství)* – olla viidennellä kuulla raskaana

*měsíc (nebeské těleso), luna* – kuu

### **5.3.4 Substantivum *týden***

Substantivum *týden* se vyskytuje hlavně v kolokacích označujících časové údaje, nebo odkazy. Kromě předložkových spojení (*před týdnem, za týden*), deiktických

odkazů (*tento týden*) jde pak podobně jako u ostatních substantiv času o spojení s adjektivy *minulý, příští* apod.

#### **TÝDEN (týdne) m – viikko**

*před týdnem* – viikko sitten, *za týden* – viikon päästä, *během týdne* – viikon aikana; *minulý týden* – viime viikolla, *příští týden* – ensi viikolla; *v týdnu* - viikolla

### **5.3.5 Substantivum *den***

Substantivum *den* se vedle typických časově odkazovacích výrazů užívá také v kolokacích označujících významné události a svátky (*Štědrý den, Mezinárodní den žen*). Tyickými spojeními s adjektivy jsou dále např. *všední den, pracovní den, soudný den*, finskému spojení *kaksi viikkoa* (*dva týdny*) odpovídá české *14 dní*.

#### **DEN (dne) m – päivä**

*jednoho dne* – eräänä päivänä, *za dva dny* – kahden päivän päästä/kuluttua; *ve dne* – päivällä, *všední den* – arkipäivä, *Štědrý den* – jouluaatto; *denně* - päivittäin, *obden* – joka toinen päivä, *14 dní* – kaksi viikkoa, *pracovní den* – työpäivä, *soudný den* – tuomiopäivä, *žít ze dne na den* – elää päivä kerrallaan

### **5.3.6 Substantivum *hodina***

Substantivum *hodina* má kromě významu časové jednotky trvání času (*2 hodiny 15 minut*), časových údajů (*Bylo 5 hodin*) a jednotky rychlosti (*50 kilometrů za hodinu*) také význam *vyučovací hodina*. Substantivum pomnožné (*hodiny, hodinky*) má pak význam přístroje na měření času (ve finštině *kello*).

#### **HODINA (hodiny) f – tunti**

*v 5 hodin* – kello 5, *viideltä, okolo třetí hodiny* - kello 3 maissa/aikaan; *kilometr za hodinu* – kilometri tunnissa; *vyučovací hodina* - oppitunti, *hodina klavíru* - pianotunti, *rozvrh hodin* - aikataulu,

*hodiny* – kello

*náramkové h.* - rannekello, *přesýpací h.*, *nástěnné h.* - seinäkello,

### 5.3.7 Substantiva *minuta*, *sekunda* a *vteřina*

Substantiva *minuta* a *sekunda* jsou substantiva původem převzatá (v souvislosti s přesnějším měřením času). Substantiva *sekunda* a *vteřina* jsou téměř absolutní synonyma. Všechna tato tři substantiva se pak užívají v podobných kontextech a kolokacích, jako jednotky času a rychlosti (8 *minut*, 10 *metrů za sekundu/vteřinu*), ve sportovním jazyce (*poslední minuty*, *závěrečné vteřiny*) apod.

**MINUTA (minuty) f – minuutti**

*v závěrečných minutách* - loppuminuuteilla; *minutku!* (*chvilku!*) – hetkinen!

**SEKUNDA (sekundy)/VTEŘINA (vteřiny) f/f – sekunti**

*poslední sekundy* – viimeiset sekunnit; *metr/megabit za sekundu* – metri/megabitti sekunnissa

## Závěr

Kontrastivní analýza korpusů finštiny a češtiny nám ukázala, že využití jednojazyčných korpusů různých jazyků a jejich vzájemná konfrontace může být jednou z možností, jak přistupovat k získávání materiálů pro výstavbu překladových vícejazyčných slovníků. Při své analýze skupiny substantiv vyjadřujících čas jsme využívali materiálů z jazykových korpusů, které měly společné rysy, ale které se také v mnohém lišily. V teoretické části práce jsme se zmínili o důležitosti parametrů užitých při výstavbě jazykových korpusů a vlivu těchto parametrů na možnosti zobecnění poznatků o jazyce získaných z korpusového materiálu. Reprezentativnost korpusu je jedním ze základních předpokladů pro využití výsledků korpusových analýz v praxi, například ve slovnících.

Rozsahem splňovaly námi použité korpusy parametry velkých korpusů (námi užitá část Kielipankki obsahovala řádově desítky milionů slov, SYN2000 i SYN2006PUB mají rozsah v řádu stamilionů slovních tvarů). Větší rozdíl byl v obsahu korpusů, SYN2000 splňuje požadavky reprezentativního korpusu psaného jazyka téměř bez výjimek, Kielipankki a SYN2006PUB jsou korpusy složené z publicistických textů, v jejich případě tedy požadavek obsahové reprezentativnosti není splněn. Otázkou v tomto případě ale zůstává, zda obsahová a tematická reprezentativnost korpusu je při použití publicistických textů, které jsou samy o sobě pestrým souborem textů různě tematicky zaměřených a různě jazykově ztvárněných, porušena natolik, že se korpus nedá považovat za uspokojivě reprezentativní pro dostatečně podložitelné a zobecnitelné analýzy. Po všech stránkách vyvážený národní jazykový korpus (těmito přívlastky můžeme určitě označit ČNK) je jistě ideální volbou pro korpusové analýzy, v případě, že takový korpus (například pro finštinu) chybí, je třeba se pokusit nahradit

ho výběrem alespoň zčásti akceptovatelného korpusu nerepresentativního. Další otázkou je také to, do jaké míry je třeba mít po všech stránkách vyvážený korpus, pokud ho chceme využívat k přípravě překladového slovníku s obsahem řádově několika tisíc hesel.

Naše korpusová analýza ukázala, že alespoň pro frekventovaná slova typu substantiv času je korpus sestavený z publicistických textů dostatečným a akceptovatelným zdrojem jazykového materiálu. K tomuto názoru nás vede jednak fakt, že jsme byli z Kielipankki schopni získat dostatek materiálu použitelného pro další zpracování, a také to, že výsledky analýz korpusových materiálů získaných z Kielipankki v mnohém potvrdily údaje uváděné v existujících slovnících vzniklých většinou na materiálech nekorpusových (ve smyslu dnešních elektronických jazykových korpusů)<sup>19</sup>. V tomto ohledu tedy jak Kielipankki, tak subkorporusy ČNK splnily podmínky pro jejich možné využití při sestavování slovníku.

Rozdílem mezi oběma použitými korpusy je kromě objemu a různé míry reprezentativnosti množství funkcí korpusového manažeru. Korpusový online manažer Lemmie 2.0 obsahuje základní statistické funkce (relativní frekvence, kolokační tabulky), korpusový manažer Bonito i webový manažer World Sketch Engine (WSE), umožňující práci s texty ČNK, obsahují kromě základních funkcí velké množství funkcí doplňujících a rozšiřujících (např. funkce Word Sketch pro možnosti sestavení sémantických polí zkoumaných slov ve WSE, či funkce Thesaurus pro vyhledání podobných a příbuzných slov a slov vyskytujících se ve stejných kontextech jako sledované slovo). Přes zdánlivou omezenost funkcí Lemmie 2.0 se ukázalo, že funkce, které Lemmie obsahuje, jsou dostačující pro zjištění základních informací nutných pro korpusový popis zkoumaných slov. Problémem ale zůstává to, že přestože Kielipankki

---

<sup>19</sup> Např.: Lindroos, Hilikka: Suomi-tšekki-suomi sanakirja, 1984, Suomen kielen perussanakirja, 1995.

obsahuje značné množství textů, nelze se všemi pracovat najednou, a maximální počet výskytů, se kterými lze pracovat v jednom okamžiku, je 20 000. S tím souvisí i další problém - statistické funkce jsou v Lemmie 2.0 počítány vzhledem k celkovému počtu slovních tvarů v celém kontextu omezeném v nastaveních korpusu, ačkoli v případě slov, jejichž absolutní frekvence v korpusu přesahuje 20 000 výskytů, nepracuje Lemmie se všemi výskyty najednou.

Substantiva vyjadřující čas jsou slova s vysokou frekvencí výskytu v jazyce a tím pádem také v jazykových korpusech. Analýza publicistických textů v Kielipankki navíc ukázala, že frekvence některých slov je v publicistických textech vyšší než v textech jiného druhu. Oproti tomu frekvenčně méně významná slova mohou mít v publicistických textech nižší frekvenci než v jazyce obecně, a publicistické texty mohou tak zkreslovat jejich obraz v jazyce. Nejen pro takovéto případy je nutné spojit výsledky získané z korpusových analýz například s materiály z existujících slovníků nebo provést doplňující analýzy jiných druhů textů, a v neposlední řadě využít možností dalších korpusových materiálů, například paralelních korpusů. Shromažďování a následné třídění a zpracovávání materiálů pro jazykové slovníky by nemělo být jednostranné (a to ani v případě využívání korpusových materiálů), ale mělo by vhodně spojovat více možností, jak uvádí například Sinclair (1991, 37-38) vedle zkoumání přirozených textů by se při sestavování slovníků měly brát v potaz názory a zkušenost uživatelů jazyka a také jiné materiály, například starší slovníky.

Srovnávací analýzy materiálů získaných z Kielipankki a Českého národního korpusu se zdají být podle našeho názoru jedním z vhodných zdrojů materiálu pro tvorbu nového překladového finsko-českého slovníku. Oba korpusy jsou snadno dostupné, práce s nimi není nijak náročná a ve spojení s vhodnými doplňujícími analyzačními postupy mohou být využity společně například s paralelními korpusy jako

základní zdroj materiálu pro nový slovník. Korpusová analýza substantiv času ve finštině a češtině nám také ukázala další možnosti pohledu na zkoumání slovní zásoby - statistickými zkoumánímí frekvencí slov, jejich tvarů a užití v různých kontextech lze dospět k novému pohledu na roli slovního tvaru ve vztahu k tvaru základnímu, k reprezentativnímu slovníkovému tvaru - lemmatu. V souvislosti s rozdílem mezi lexémem a jeho realizacemi v jazyce zmiňuje Sinclair (1998) důležitost přehodnocení lexikografického pohledu na jazyk, novou jednotkou slovní zásoby by měla podle něho být lexikální jednotka, vymezená formálně a sémanticky a obsahující všechny aspekty fungující jazykové jednotky. Lexikální jednotka by pak nebyla omezena na pouhý základní tvar slova reprezentující všechny jeho možné varianty výskytu, ale byla by živou a kontextově zakotvenou jednotkou dávající uživateli slovníku možnost seznámit se s jazykem blíže jeho konkrétní realizaci.



## **Přílohy:**

1. Kolokační tabulky substantiv času ve finštině (Kielipankki)
2. Kolokační tabulky substantiv času v češtině (ČNK)

## **Příloha č. 1: Kolokační tabulky substantiv času ve finštině**

V příloze 1 uvádíme kolokační tabulky substantiv času ve finštině. Pro každé ze zkoumaných substantiv jsme vygenerovali kolokační tabulky pro obě statistické hodnoty míry asociace (MI score i t-score). Kolokační tabulka obsahuje v titulku informace o konkrétním substantivu, název statistické hodnoty, podle níž je řazena a kontext, v němž byly kolokace počítány (-3/+3 znamená kontext 3 slov nalevo i napravo od nodu – konkrétního výskytu zkoumaného slova). První a třetí sloupec uvádí kolokáty – slova/slovní tvary, s nimiž daný tvar zkoumaného substantiva tvoří kolokaci. Další sloupce uvádějí statistické hodnoty t-score/MI score (v závislosti na statistické hodnotě, podle níž byla tabulka sestavena), informace o celkovém počtu kolokátů a jejich pozici vůči danému nodu, relativní frekvenci kolokace a informace o celkovém počtu výskytu obou členů kolokace v korpusu.

## Substantivum aika

Kolokačnī tabulka "aika" t-score (-3/+3)								
	Kolokace			t-score	Abs (Nalevo/Napravo)	Rel	Celkem nodū	Celkem kolokātū
1	vuoden	ajan		28,82954703	906 ( 897 / 9 )	4,1062E-05	5140	27368
2	kaikkien	aikojen		27,70905621	770 ( 770 / 0 )	3,4898E-05	1165	3530
3		aika	on	24,7633462	1291 ( 542 / 749 )	5,8511E-05	3647	404609
4	koko	ajan		19,76010657	433 ( 429 / 4 )	1,9625E-05	5140	15613
5	suomen	aikaa		18,98900924	429 ( 426 / 3 )	1,9443E-05	3598	36485
6		aika	sulkuihin	18,66289663	349 ( 0 / 349 )	1,5817E-05	3647	354
7		aika	merkitty	18,64362603	349 ( 0 / 349 )	1,5817E-05	3647	717
8	paikallinen	aika		18,64288281	349 ( 349 / 0 )	1,5817E-05	3647	731
9	sama	aika		16,60533684	282 ( 278 / 4 )	1,2781E-05	3647	3177
10	viikon	ajan		14,2168247	221 ( 218 / 3 )	1,0016E-05	5140	6906
11	viime	aikojen		12,31861098	168 ( 164 / 4 )	7,6141E-06	1165	26304
12	parin	ajan		12,25536731	156 ( 154 / 2 )	7,0703E-06	5140	2097
13	kolmen	ajan		11,821096	154 ( 152 / 2 )	6,9796E-06	5140	5226
14	jonka	aikana		11,56584893	143 ( 142 / 1 )	6,4811E-06	1021	16903
15	lyhyessä	ajassa		11,17754334	125 ( 125 / 0 )	5,6653E-06	570	204
16	välisenä	aikana		11,12733206	124 ( 124 / 0 )	0,00000562	1021	330
17	kuukauden	ajan		10,94483912	128 ( 124 / 4 )	5,8012E-06	5140	2986
18	tunnin	ajan		10,93922948	126 ( 126 / 0 )	5,7106E-06	5140	2295
19	paljon	aikaa		10,18985163	130 ( 117 / 13 )	5,8919E-06	3598	14123
20	vastaavana	aikana		10,14336176	103 ( 103 / 0 )	4,6682E-06	1021	203
21	on	aikaa		9,999260941	651 ( 529 / 122 )	2,9505E-05	3598	404609
22	muutosten	aika		9,867458974	98 ( 97 / 1 )	4,4416E-06	3647	320
23	olisi	aika		9,661321547	140 ( 119 / 21 )	6,3451E-06	3647	25900
24	muutaman	ajan		9,600548799	100 ( 100 / 0 )	4,5322E-06	5140	2858
25		ajat	ovat	9,529927088	107 ( 27 / 80 )	4,8495E-06	414	74808
26	sinä	aikana		9,521453885	91 ( 91 / 0 )	4,1243E-06	1021	617
27	vuosien	ajan		9,412094628	94 ( 94 / 0 )	4,2603E-06	5140	1965
28	pitkän	ajan		9,364644493	92 ( 92 / 0 )	4,1696E-06	5140	1558
29	muina	aikoina		9,216441019	85 ( 85 / 0 )	3,8524E-06	481	220
30	vie	aikaa		9,190294819	88 ( 85 / 3 )	3,9884E-06	3598	1827
31	kello	aikaa		9,127558583	93 ( 84 / 9 )	4,215E-06	3598	5087
32		aika	ohi	9,099753708	87 ( 1 / 86 )	3,943E-06	3647	2141
33	päivän	ajan		9,07499886	100 ( 100 / 0 )	4,5322E-06	5140	6618
34		ajan	kuluessa	8,671194467	79 ( 0 / 79 )	3,5805E-06	5140	1380
35		aikana	jolloin	8,494168049	75 ( 0 / 75 )	3,3992E-06	1021	5181
36	näinä	aikoina		8,481410656	72 ( 72 / 0 )	3,2632E-06	481	252
37	enemmän	aikaa		8,477890155	90 ( 75 / 15 )	4,079E-06	3598	9783
38	tämän	ajan		8,44661646	94 ( 89 / 5 )	4,2603E-06	5140	8662
39	tuohon	aikaan		8,241598206	68 ( 68 / 0 )	3,0819E-06	691	203
41	korkea	aika		8,188797297	69 ( 69 / 0 )	3,1272E-06	3647	987
42	lyhyen	ajan		8,169101162	68 ( 67 / 1 )	3,0819E-06	5140	455
43	pitkä	aika		8,074300709	69 ( 65 / 4 )	3,1272E-06	3647	1946
44	viime	aikana		8,004857335	78 ( 65 / 13 )	3,5351E-06	1021	26304
45	tänä	aikana		7,93659867	67 ( 67 / 0 )	3,0366E-06	1021	7334
46		aika	paikka	7,768690432	66 ( 1 / 65 )	2,9913E-06	3647	2911

Kolokačni tabulka "aika" MI score (-3/+3)									
	Kolokace			MI score	Abs (Nalevo/Napravo)	Rel	Celkem nodů	Celkem kolokátů	
1	tyrmäystippoja	ajasta		13,16829363	72 ( 72 / 0 )	3,2632E-06	394	73	
2		noihin	aikoihin	13,06084482	25 ( 25 / 0 )	1,1331E-06	203	53	
3	esihistoriallisella	ajalla		12,87704615	6 ( 6 / 0 )	2,719E-07	419	7	
4		ankeina	aikoina	12,63731751	5 ( 5 / 0 )	2,266E-07	481	6	
5		edeltävältä	ajalta	12,61745904	5 ( 5 / 0 )	2,266E-07	418	7	
6		iät	ajat	12,59738389	30 ( 30 / 0 )	1,3597E-06	414	43	
7		öiseen	aikaan	12,0297798	11 ( 11 / 0 )	4,985E-07	691	14	
8		lyhyessä	ajassa	11,94878584	125 ( 125 / 0 )	5,6653E-06	570	204	
9		vaikeina	aikoina	11,9409939	18 ( 18 / 0 )	8,158E-07	481	35	
10		tietyksi	ajaksi	11,87507995	6 ( 6 / 0 )	2,719E-07	534	11	
11		nykyisinä	aikoina	11,63731751	5 ( 5 / 0 )	2,266E-07	481	12	
12		muina	aikoina	11,52838314	85 ( 85 / 0 )	3,8524E-06	481	220	
13		entisinä	aikoina	11,50180254	11 ( 10 / 1 )	4,985E-07	481	29	
14		vastaavasta	ajasta	11,47878331	37 ( 37 / 0 )	1,6769E-06	394	121	
15		huonoina	aikoina	11,41002629	21 ( 21 / 0 )	9,518E-07	481	59	
16		havaintoja	ajasta	11,37512461	72 ( 72 / 0 )	3,2632E-06	394	253	
17		sodanjälkeisenä	aikana	11,35504623	8 ( 8 / 0 )	3,626E-07	1021	11	
18		näinä	aikoina	11,092997	72 ( 72 / 0 )	3,2632E-06	481	252	
19		kaikkina	aikoina	11,01582914	26 ( 26 / 0 )	1,1784E-06	481	96	
20		joksikin	ajaksi	10,94219415	6 ( 6 / 0 )	2,719E-07	534	21	
21		vanhoina	aikoina	10,93351878	11 ( 11 / 0 )	4,985E-07	481	43	
22		tuolta	ajalta	10,91024079	14 ( 14 / 0 )	6,345E-07	418	64	
23		vastaavana	aikana	10,83564246	103 ( 103 / 0 )	4,6682E-06	1021	203	
24		tuohon	aikaan	10,79983003	68 ( 68 / 0 )	3,0819E-06	691	203	
25		tuona	aikana	10,79779911	43 ( 43 / 0 )	1,9489E-06	1021	87	
26		muuna	aikana	10,71494218	7 ( 7 / 0 )	3,173E-07	1021	15	
27		hyvinä	aikoina	10,68403401	34 ( 34 / 0 )	1,541E-06	481	158	
28		äitiysloman	ajalta	10,43992086	6 ( 6 / 0 )	2,719E-07	418	38	
29		välisenä	aikana	10,40235195	124 ( 124 / 0 )	0,00000562	1021	330	
30		minun	aikanani	10,30229793	13 ( 13 / 0 )	5,892E-07	20	1893	
31		tiettyä	aikana	10,17062166	8 ( 8 / 0 )	3,626E-07	1021	25	
32		myöhempien	aikojen	10,07981025	12 ( 12 / 0 )	5,439E-07	1165	35	
33		hyviin	aikoihin	10,02943187	6 ( 6 / 0 )	2,719E-07	203	104	
34		niinä	aikoina	10,0258828	6 ( 6 / 0 )	2,719E-07	481	44	
35		omaan	aikaamme	9,991662825	8 ( 8 / 0 )	3,626E-07	24	1204	
36		väärään	aikaan	9,976823666	32 ( 31 / 1 )	1,4503E-06	691	169	
37			aika	sulkuihin	9,957228195	349 ( 0 / 349 )	1,5817E-05	3647	354
38		väliseksi	ajaksi	9,942194149	8 ( 8 / 0 )	3,626E-07	534	56	
39		viettivät	aikaansa	9,911265798	5 ( 5 / 0 )	2,266E-07	201	95	
41		parempia	aikoja	9,820407047	24 ( 24 / 0 )	1,0877E-06	261	374	
42		jälkeisenä	aikana	9,814477852	10 ( 10 / 0 )	4,532E-07	1021	40	
43		meidän	aikanamme	9,724689556	12 ( 12 / 0 )	5,439E-07	15	3477	
44		menneinä	aikoina	9,652424406	6 ( 6 / 0 )	2,719E-07	481	57	
45		sovittuun	aikaan	9,529706196	5 ( 5 / 0 )	2,266E-07	691	36	

Substantivum *vuosi*

Kolokačnǐ tabulka "vuosi" (t-score -3/+3)								
	Kolokace			t-score	Abs (Nalevo/Napravo)	Rel	Celkem nodů	Celkem kolokátů
1		vuotta	sitten	38,1350255	1491 ( 3 / 1488 )	6,7575E-05	4439	18415
2	viime	vuonna		30,68102702	1491 ( 3 / 1488 )	4,4869E-05	4138	26304
3	tänä	vuonna		30,59960267	990 ( 989 / 1 )	4,3056E-05	4138	7334
4	ensi	vuoden		25,97563254	950 ( 949 / 1 )	3,2224E-05	5619	14436
5	ensi	vuonna		25,81107546	711 ( 705 / 6 )	3,1408E-05	4138	14436
6	50	vuotta		22,15149684	693 ( 689 / 4 )	2,3613E-05	4439	15298
7	tämän	vuoden		21,75309782	521 ( 520 / 1 )	2,2435E-05	5619	8662
8		vuoden	aikana	21,6759295	495 ( 493 / 2 )	2,2706E-05	5619	12434
9	viime	vuoden		20,15944669	501 ( 3 / 498 )	2,1347E-05	5619	26304
10		vuosi	sitten	19,39147353	471 ( 471 / 0 )	1,7313E-05	719	18415
11		vuoden	alusta	17,96816021	382 ( 2 / 380 )	1,4956E-05	5619	2823
12	viime	vuosina		16,90666969	330 ( 45 / 285 )	1,3325E-05	690	26304
13	kaksi	vuotta		16,81956467	294 ( 294 / 0 )	1,4367E-05	4439	17435
14		vuoden	ajan	16,35863171	317 ( 313 / 4 )	1,2781E-05	5619	5728
15	kymmenen	vuotta		16,17313313	282 ( 2 / 280 )	1,2554E-05	4439	7781
16		vuoden	alussa	16,03005277	277 ( 276 / 1 )	1,2328E-05	5619	5990
17		vuoden	loppuun	15,70032958	272 ( 5 / 267 )	1,1557E-05	5619	3367
18		vuoden	kuluttua	15,53994045	255 ( 2 / 253 )	1,1195E-05	5619	2177
19	pari	vuotta		15,50845252	247 ( 2 / 245 )	1,1331E-05	4439	4763
20		vuodesta	lähtien	12,98930086	199 ( 196 / 3 )	7,7048E-06	950	2977
21		vuotta	myöhemmin	11,9681857	161 ( 157 / 4 )	6,9343E-06	4439	4933
22	muutama	vuosi		11,88042604	155 ( 155 / 0 )	6,4357E-06	719	2632
23	markkaa	vuodessa		11,76118956	145 ( 137 / 8 )	6,6624E-06	903	21520
24	puoli	vuotta		10,95158809	155 ( 154 / 1 )	5,7106E-06	4439	3051
25	sata	vuotta		10,89093336	126 ( 126 / 0 )	5,5746E-06	4439	2201
26		vuoden	lopussa	10,52017684	131 ( 130 / 1 )	5,348E-06	5619	2923
27		vuoden	jälkeen	10,50348163	118 ( 4 / 114 )	7,7954E-06	5619	26897
28	muutaman	vuoden		10,48024045	172 ( 8 / 164 )	5,3027E-06	5619	2858
29	joka	vuosi		10,09621427	108 ( 108 / 0 )	5,2121E-06	719	41308
30	jo	vuonna		9,951948749	105 ( 0 / 105 )	7,2062E-06	4138	35737
31		vuoden	lopulla	9,663873842	102 ( 2 / 100 )	4,4869E-06	5619	2235
32	samana	vuonna		8,746729817	86 ( 68 / 18 )	3,5351E-06	4138	801
33		vuoden	päästä	8,677102155	78 ( 75 / 3 )	3,8071E-06	5619	3513
34		vuoteen	mennessä	8,671415724	84 ( 5 / 79 )	3,4445E-06	562	3176
35	seuraavana	vuonna		8,574295795	80 ( 78 / 2 )	3,3992E-06	4138	794
36		vuoden	mennessä	8,132223522	67 ( 56 / 11 )	3,3538E-06	5619	3176
37	useita	vuosia		8,08978567	106 ( 101 / 5 )	2,9913E-06	370	3320
38	täyttää	vuotta		8,073777828	66 ( 66 / 0 )	3,0366E-06	4439	908
39		vuoden	kuluessa	7,650855172	125 ( 109 / 16 )	0,00000281	5619	1380
40		vuonna	syntynyt	7,625608601	62 ( 1 / 61 )	2,8553E-06	4138	2638

Kolokačnī tabulka "vuosi" MI score (-3/+3)								
	Kolokace			MI score	Abs (Nal./Napr.)	Rel	Celkem nodū	Celkem kolokātū
1	juhlii	vuottaan		12,23206103	5 ( 5 / 0 )	2,266E-07	13	294
2	kolmena	vuotena		11,67479222	5 ( 5 / 0 )	2,266E-07	37	152
3		vuotensa	vietti	11,62678685	7 ( 2 / 5 )	3,173E-07	20	407
4	kymmeneksi	vuodeksi		11,33515643	22 ( 20 / 2 )	9,971E-07	307	102
5	viideksi	vuodeksi		11,25864354	27 ( 27 / 0 )	1,2237E-06	307	132
6	moneksi	vuodeksi		10,96318766	8 ( 8 / 0 )	3,626E-07	307	48
7	puoleksi	vuodeksi		10,74079524	19 ( 19 / 0 )	8,611E-07	307	133
8	viimeiset	vuotensa		10,73009369	12 ( 12 / 0 )	5,439E-07	20	1299
9	tälle	vuodelle		10,55760645	22 ( 22 / 0 )	9,971E-07	87	617
10	aikaisempina	vuosina		10,30940312	10 ( 10 / 0 )	4,532E-07	690	42
11	pyöreitä	vuosia		10,17452688	5 ( 5 / 0 )	2,266E-07	370	43
12	tulevina	vuosina		10,01722237	21 ( 21 / 0 )	9,518E-07	690	108
13	noina	vuosina		10,01055864	6 ( 6 / 0 )	2,719E-07	690	31
14	kuudeksi	vuodeksi		9,963187657	5 ( 3 / 2 )	2,266E-07	307	60
15	kahdeksi	vuodeksi		9,948832364	25 ( 24 / 1 )	1,1331E-06	307	303
16	täksi	vuodeksi		9,874061766	17 ( 17 / 0 )	7,705E-07	307	217
17	kolmelle	vuodelle		9,862681115	5 ( 5 / 0 )	2,266E-07	87	227
18		vuosien	saatossa	9,76285597	9 ( 0 / 9 )	4,079E-07	401	95
19	pariksi	vuodeksi		9,706847904	9 ( 9 / 0 )	4,079E-07	307	129
20	omakuva	vuodelta		9,591580868	6 ( 6 / 0 )	2,719E-07	362	79
21	edelliseen	vuoteen		9,495246437	15 ( 15 / 0 )	6,798E-07	562	136
22		vuosien	mittaan	9,35905197	57 ( 0 / 57 )	2,5834E-06	401	796
23		vuodelta	huudettiin	9,284864024	7 ( 0 / 7 )	3,173E-07	362	114
24	kuluva	vuosi		9,204919825	6 ( 6 / 0 )	2,719E-07	719	52
25	aiempina	vuosina		9,17815859	5 ( 5 / 0 )	2,266E-07	690	46
26	kwh	vuodessa		9,104137555	5 ( 4 / 1 )	2,266E-07	903	37
27		vuodeksi	vankeuteen	9,028282685	17 ( 1 / 16 )	7,705E-07	307	390
28	aikaisempien	vuosien		9,00898124	5 ( 5 / 0 )	2,266E-07	401	89
29		vuoden	kyläksi	8,939109819	6 ( 0 / 6 )	2,719E-07	5619	8
30	aiempien	vuosien		8,889768082	6 ( 6 / 0 )	2,719E-07	401	116
31		vuosien	varrella	8,840858482	43 ( 0 / 43 )	1,9489E-06	401	860
32	menneiden	vuosien		8,729827169	5 ( 5 / 0 )	2,266E-07	401	108
33	vajaassa	vuodessa		8,690493291	7 ( 7 / 0 )	3,173E-07	903	69
34		vuosiksi	eteenpäin	8,675818683	5 ( 0 / 5 )	2,266E-07	32	1405
35	tähän aikaan	vuodesta		8,572686463	6 ( 6 / 0 )	2,719E-07	950	61
36	teoksia	vuosilta		8,496738376	5 ( 5 / 0 )	2,266E-07	62	821
37	seuraavina	vuosina		8,389846116	9 ( 9 / 0 )	4,079E-07	690	143
38	alaikäraja	vuotta		8,27918851	6 ( 6 / 0 )	2,719E-07	4439	16
39	muutama	vuosi		8,118324605	143 ( 142 / 1 )	6,4811E-06	719	2632
40	joka toinen	vuosi		8,110943677	18 ( 18 / 0 )	8,158E-07	719	333

Substantivum *kuukausi*

Kolokačnı tabulka "kuukausi" t-score (-3/+3)								
	Kolokace			t-score	Abs (Nal./Napr.)	Rel.	Celkem nodů	Celkem kolokátů
1	markkaa	kuukaudessa		23,90146477	586 ( 551 / 35 )	2,6559E-05	1267	21520
2	kolmen	kuukauden		21,01337857	450 ( 443 / 7 )	2,0395E-05	2986	5226
3		kuukauden	aikana	17,94743515	342 ( 6 / 336 )	1,55E-05	2986	12434
4		kuukausi	sitten	16,68303186	284 ( 0 / 284 )	1,2872E-05	570	18415
5		kuukauden	kuluttua	15,18158042	234 ( 0 / 234 )	1,0605E-05	2986	2177
6	kolme	kuukautta		14,58797602	224 ( 222 / 2 )	1,0152E-05	1658	12572
7		kuukauden	korko	14,18263705	207 ( 3 / 204 )	9,3817E-06	2986	3631
8	yhdeksän	kuukautta		14,13352338	202 ( 201 / 1 )	9,1551E-06	1658	2497
9		kuukautta	sitten	13,21957135	191 ( 2 / 189 )	8,6565E-06	1658	18415
10	viime	kuukausina		13,07979381	174 ( 173 / 1 )	7,8861E-06	205	26304
11	pari	kuukautta		12,95118446	172 ( 171 / 1 )	7,7954E-06	1658	4763
12	muutaman	kuukauden		12,50572846	161 ( 161 / 0 )	7,2969E-06	2986	2858
13	neljä	kuukautta		11,62560185	141 ( 136 / 5 )	6,3904E-06	1658	6552
14		kuukauden	ajan	11,56935445	143 ( 4 / 139 )	6,4811E-06	2986	5728
15	parin	kuukauden		10,8912316	122 ( 118 / 4 )	5,5293E-06	2986	2097
16	ensimmäisen	kuukauden		10,79254341	125 ( 123 / 2 )	5,6653E-06	2986	5340
17		kuukauden	jälkeen	10,08362276	142 ( 9 / 133 )	6,4357E-06	2986	26897
18	kahdeksan	kuukauden		10,04453966	106 ( 105 / 1 )	4,8042E-06	2986	3184
19		kuukauden	kuluessa	9,938399555	101 ( 0 / 101 )	4,5775E-06	2986	1380
20		kuukautta	myöhemmin	8,752891925	81 ( 0 / 81 )	3,6711E-06	1658	4933
21		kuukaudeksi	vankeuteen	8,598986331	74 ( 3 / 71 )	3,3538E-06	272	390
22	muutama	kuukausi		8,377746728	71 ( 71 / 0 )	3,2179E-06	570	2632
23	viisi	kuukautta		7,691506747	64 ( 61 / 3 )	2,9006E-06	1658	5474
24	kerran	kuukaudessa		7,555018102	61 ( 59 / 2 )	2,7647E-06	1267	5786
25		kuukauden	sisällä	7,351089699	57 ( 0 / 57 )	2,5834E-06	2986	1848
26	useita	kuukausia		7,17855945	52 ( 51 / 1 )	2,3568E-06	260	3320
27	jo	kuukauden		7,033454895	99 ( 84 / 15 )	4,4869E-06	2986	35737
28	joka	kuukausi		6,990462673	61 ( 61 / 0 )	2,7647E-06	570	41308
29	kestää	kuukautta		6,963636823	50 ( 50 / 0 )	2,2661E-06	1658	1685
30	muutamassa	kuukaudessa		6,755311672	46 ( 46 / 0 )	2,0848E-06	1267	532
31		kuukautta	ennen	6,708289516	57 ( 1 / 56 )	2,5834E-06	1658	14092
32	yksi	kuukausi		6,673901251	48 ( 47 / 1 )	2,1755E-06	570	11367
33	noin	kuukauden		6,639068609	76 ( 66 / 10 )	3,4445E-06	2986	22318
34		kuukauden	korot	6,588746838	46 ( 3 / 43 )	2,0848E-06	2986	1617
35		kuukauden	päästä	6,516483292	48 ( 1 / 47 )	2,1755E-06	2986	3513
36	viime	kuukausien		6,487679552	44 ( 44 / 0 )	1,9942E-06	135	26304
37	12	kuukauden		6,295201912	62 ( 61 / 1 )	0,00000281	2986	15310
38	säilytetään	kuukausi		6,240411784	39 ( 39 / 0 )	1,7676E-06	570	185
39		kuukautta	vankeutta	6,065349003	37 ( 3 / 34 )	1,6769E-06	1658	235

Kolokační tabulka "kuukausi" MI score (-3/+3)								
	Kolokace			MI score	Abs (Nal./Napr.)	Rel	Celkem nodů	Celkem kolokátů
1	perättäisenä	kuukautena		14,27108619	5 ( 5 / 0 )	2,266E-07	31	30
2	irtisanomisaika	kuukaudesta		13,41302878	7 ( 7 / 0 )	3,173E-07	59	40
3	ensimmäiseltä	kuukaudelta		13,19296122	10 ( 10 / 0 )	4,532E-07	51	77
4	kuudeksi	kuukaudeksi		12,81589157	32 ( 32 / 0 )	1,4503E-06	272	60
5	yhdeksäksi	kuukaudeksi		12,58527864	10 ( 10 / 0 )	4,532E-07	272	22
6	muutamalla	kuukaudella		11,59390332	6 ( 6 / 0 )	2,719E-07	43	166
7		kuukaudeksi	ehdolliseen	11,57839225	19 ( 0 / 19 )	8,611E-07	272	84
8	viikosta	kuukauteen		11,46468135	7 ( 7 / 0 )	3,173E-07	132	69
9		kuukaudeksi	vankeuteen	11,32490521	74 ( 3 / 71 )	3,3538E-06	272	390
10	kolmeksi	kuukaudeksi		11,03136872	37 ( 37 / 0 )	1,6769E-06	272	239
11		kuukaudeksi	ehdottomaan	10,57303504	8 ( 0 / 8 )	3,626E-07	272	71
12	säilytetään	kuukausi		10,40944765	39 ( 39 / 0 )	1,7676E-06	570	185
13	pariksi	kuukaudeksi		10,17098652	11 ( 9 / 2 )	4,985E-07	272	129
14	muutamaksi	kuukaudeksi		10,15292655	8 ( 8 / 0 )	3,626E-07	272	95
15	edellisestä	kuukaudesta		10,07460437	5 ( 5 / 0 )	2,266E-07	59	289
16	kymmeneksi	kuukaudeksi		10,05035682	8 ( 8 / 0 )	3,626E-07	272	102
17	kuudesta	kuukaudesta		10,01591037	5 ( 5 / 0 )	2,266E-07	59	301
18		kuukauden	heliborkorko	9,796761269	13 ( 0 / 13 )	5,892E-07	2986	18
19	viimeisinä	kuukausina		9,684508673	5 ( 5 / 0 )	2,266E-07	205	109
20		kuukauden	julkaisusta	9,559977756	38 ( 0 / 38 )	1,7222E-06	2986	62
21	kahdesta	kuukaudesta		9,473645465	9 ( 9 / 0 )	4,079E-07	59	789
22	muutamien	kuukausien		9,397146018	5 ( 5 / 0 )	2,266E-07	135	202
23	nykyisestä	kuukaudesta		9,342639453	6 ( 6 / 0 )	2,719E-07	59	576
24	vuodeksi	kuukaudeksi		9,248618299	62 ( 60 / 2 )	0,00000281	272	1378
25		kuukaudeksi	vankilaan	9,222211848	11 ( 4 / 7 )	4,985E-07	272	249
26		kuukaudesta	vuoteen	9,20951437	25 ( 0 / 25 )	1,1331E-06	59	2632
27		kuukauden	tähtitaivas	9,183784393	17 ( 0 / 17 )	7,705E-07	2986	36
28	kolmesta	kuukauteen		9,169246976	18 ( 18 / 0 )	8,158E-07	132	871
29	kolmeen	kuukauteen		9,141360019	15 ( 15 / 0 )	6,798E-07	132	740



Substantivum *viikko*

Kolokačni tabulka "viikko" t-score (-3/+3)								
	Kolokace			t-score	Abs (Nal./Napr.)	Rel.	Celkem nodů	Celkem kolokátů
1	viime	viikolla		37,20172398	1440 ( 1435 / 5 )	6,5264E-05	3962	26304
2	tällä	viikolla		30,94066049	973 ( 972 / 1 )	4,4099E-05	3962	7324
3	ensi	viikolla		30,82502518	981 ( 971 / 10 )	4,4461E-05	3962	14436
4		viikkoa	sitten	28,48463247	835 ( 4 / 831 )	3,7844E-05	2379	18415
5		viikko	sitten	27,83266543	793 ( 3 / 790 )	3,5941E-05	1845	18415
6	pari	viikkoa		23,44927054	556 ( 556 / 0 )	2,5199E-05	2379	4763
7	kaksi	viikkoa		23,07995273	555 ( 549 / 6 )	2,5154E-05	2379	17435
8		viikon	kuluttua	22,44699662	512 ( 1 / 511 )	2,3205E-05	6906	2177
9	ensi	viikon		20,15909114	459 ( 454 / 5 )	2,0803E-05	6906	14436
10	viime	viikon		18,91088728	451 ( 449 / 2 )	2,044E-05	6906	26304
11	parin	viikon		17,92137613	329 ( 320 / 9 )	1,4911E-05	6906	2097
12		viikon	aikana	17,0856978	337 ( 7 / 330 )	1,5274E-05	6906	12434
13	kolme	viikkoa		16,46109445	287 ( 284 / 3 )	1,3008E-05	2379	12572
14		viikkoa	ennen	16,15760343	279 ( 1 / 278 )	1,2645E-05	2379	14092
15		viikon	päästä	15,32917242	248 ( 5 / 243 )	1,124E-05	6906	3513
16		viikon	ajan	15,22977353	253 ( 3 / 250 )	1,1467E-05	6906	5728
17	tämän	viikon		14,95012834	255 ( 255 / 0 )	1,1557E-05	6906	8662
18		viikon	runo	14,94436309	225 ( 0 / 225 )	1,0198E-05	6906	445
19	viime	viikkoina		14,12236254	203 ( 203 / 0 )	9,2004E-06	250	26304
20	muutama	viikko		13,24274048	178 ( 176 / 2 )	8,0673E-06	1845	2632
21	kerran	viikossa		13,02846919	173 ( 171 / 2 )	7,8407E-06	1041	5786
22	edellisen	viikon		12,38601686	159 ( 159 / 0 )	7,2062E-06	6906	1501
23	muutaman	viikon		12,34679643	163 ( 163 / 0 )	7,3875E-06	6906	2858
24	kertaa	viikossa		11,99977208	148 ( 145 / 3 )	6,7077E-06	1041	7125
25	viimeistään	viikkoa		10,79795438	118 ( 118 / 0 )	5,348E-06	2379	1089
26		viikon	sisällä	10,73118744	122 ( 1 / 121 )	5,5293E-06	6906	1848
27	jo	viikolla		10,68156966	183 ( 160 / 23 )	8,294E-06	3962	35737
28		viikon	lopulla	10,47641916	118 ( 1 / 117 )	5,348E-06	6906	2235
29		viikkoa	myöhemmin	10,4740524	116 ( 1 / 115 )	5,2574E-06	2379	4933
30		viikon	alussa	10,17422404	125 ( 1 / 124 )	5,6653E-06	6906	5990
31	runsas	viikko		10,11460687	103 ( 103 / 0 )	4,6682E-06	1845	694
32		viikon	kuluessa	9,842921212	102 ( 0 / 102 )	4,6229E-06	6906	1380
33	runsaan	viikon		9,379360329	91 ( 90 / 1 )	4,1243E-06	6906	813
34	tuntia	viikossa		9,362854207	89 ( 80 / 9 )	4,0337E-06	1041	2371
35	kuluneella	viikolla		9,261205401	86 ( 86 / 0 )	3,8977E-06	3962	107
36	jo	viikkoa		8,734500285	118 ( 113 / 5 )	5,348E-06	2379	35737
37		viikon	levy	8,637154103	77 ( 0 / 77 )	3,4898E-06	6906	644
38		viikkoa	julkaisupäivää	8,596774591	74 ( 0 / 74 )	3,3538E-06	2379	74
39	päivänä	viikossa		8,390648281	72 ( 71 / 1 )	3,2632E-06	1041	2837
40		viikon	jälkeen	8,123208798	150 ( 26 / 124 )	6,7983E-06	6906	26897

Kolokačni tabulka "viikko" MI score (-3/+3)								
	Kolokace			MI score	Abs (Nal./Napr.)	Rel.	Celkem nodú	Celkem kolokátú
1	edeltävinä	viikkoina		12,93757012	8 ( 8 / 0 )	3,626E-07	250	15
2	muistoja	viikolta		12,39661707	13 ( 13 / 0 )	5,892E-07	31	286
3		viikosta	viikkoon	11,6719732	9 ( 0 / 9 )	4,079E-07	69	147
4	pariksi	viikoksi		11,42050612	20 ( 18 / 2 )	9,064E-07	208	129
5	kahdeksi	viikoksi		11,22418331	41 ( 37 / 4 )	1,8582E-06	208	303
6		viikkojen	lukumäärä	10,95849596	40 ( 0 / 40 )	1,8129E-06	220	336
7	muutamaksi	viikoksi		10,70987468	9 ( 9 / 0 )	4,079E-07	208	95
8		viikkoa	julkaisupäivää	10,59410545	74 ( 0 / 74 )	3,3538E-06	2379	74
9	kolmeksi	viikoksi		10,53086657	20 ( 20 / 0 )	9,064E-07	208	239
10	seitsemäksi	viikoksi		10,52484278	5 ( 5 / 0 )	2,266E-07	208	60
11		viikkoa	merkkipäivää	10,00009781	53 ( 0 / 53 )	2,4021E-06	2379	80
12		viikkojen	kuukausien	9,759424614	7 ( 0 / 7 )	3,173E-07	220	135
13	kirjallisena	viikkoa		9,704288364	34 ( 34 / 0 )	1,541E-06	2379	63
14	toimituksessa	viikkoa		9,682642121	126 ( 126 / 0 )	5,7106E-06	2379	237
15		viikkoa	toivottua	9,627788573	65 ( 0 / 65 )	2,9459E-06	2379	127
16	tälle	viikolle		9,571491973	6 ( 6 / 0 )	2,719E-07	47	617
17	kuluneella	viikolla		9,543029606	86 ( 86 / 0 )	3,8977E-06	3962	107
18	muutamien	viikkojen		9,540598808	9 ( 9 / 0 )	4,079E-07	220	202
19	muutamaan	viikkoon		9,408938797	5 ( 5 / 0 )	2,266E-07	147	184
20	viideksi	viikoksi		9,38733926	5 ( 5 / 0 )	2,266E-07	208	132
21	päivästä	viikkoon		9,298986507	7 ( 7 / 0 )	3,173E-07	147	278
22		viikon	julkisjuorut	9,056610384	13 ( 0 / 13 )	5,892E-07	6906	13
23		viikon	puhuttavimmat	9,056610384	7 ( 0 / 7 )	3,173E-07	6906	7
24		viikon	luontoääni	9,043063852	106 ( 0 / 106 )	4,8042E-06	6906	107
25	näinä	viikkoina		9,037105796	9 ( 9 / 0 )	4,079E-07	250	252
26	jälkeisellä	viikolla		9,000250843	16 ( 16 / 0 )	7,252E-07	3962	29
27		viikon	ensi-iltaelokuvat	8,95473077	41 ( 0 / 41 )	1,8582E-06	6906	44
28	kolme-neljä	viikossa		8,860491231	5 ( 5 / 0 )	2,266E-07	1041	38
29		viikon	avainteko	8,754047614	30 ( 0 / 30 )	1,3597E-06	6906	37
30	neljänä	viikossa		8,671013432	9 ( 9 / 0 )	4,079E-07	1041	78
31	kahdesta	viikosta		8,662801557	6 ( 6 / 0 )	2,719E-07	69	789
32		viikon	kasvo	8,581701429	59 ( 0 / 59 )	2,674E-06	6906	82
33	muutamia	viikkoja		8,57713717	24 ( 24 / 0 )	1,0877E-06	227	1018

Substantivum päivä

Kolokačni tabulka "päivä" t-score (-3)							
	Kolokace		t-score	Abs (Nal./Napr.)	Rel.	Celkem nodû	Celkem kolokátû
1	viime	päivinä	16,87471459	290 ( 290 / 0 )	1,31434E-05	737	26304
2	onnen	päivät	16,81915249	283 ( 283 / 0 )	1,28262E-05	687	641
3	samana	päivänä	16,32184429	267 ( 267 / 0 )	0,000012101	2759	801
4	aikaisen	päivä	15,84244083	252 ( 252 / 0 )	1,14212E-05	2101	1787
5	ysten	päivä	15,76914187	251 ( 251 / 0 )	1,13759E-05	2101	4100
6	seuraavana	päivänä	15,76094701	249 ( 249 / 0 )	1,12852E-05	2759	794
7	pari	päivää	14,03234219	200 ( 200 / 0 )	9,0644E-06	2399	4763
8	tänä	päivänä	13,69446704	193 ( 193 / 0 )	8,7472E-06	2759	7334
9	koko	päivän	13,68672034	214 ( 214 / 0 )	9,6989E-06	6492	15613
10	7	päivää	13,30705976	186 ( 186 / 0 )	8,4299E-06	2399	13847
11	kolme	päivää	13,22466558	183 ( 183 / 0 )	0,000008294	2399	12572
12	muutaman	päivän	12,6488847	165 ( 165 / 0 )	7,4782E-06	6492	2858
13	kaksi	päivää	12,4427477	166 ( 166 / 0 )	7,5235E-06	2399	17435
14	lämpötila	päivällä	12,2174716	150 ( 150 / 0 )	6,7983E-06	1573	1719
15	kolmen	päivän	12,03799251	154 ( 154 / 0 )	6,9796E-06	6492	5226
16	uutiset	päivät	11,57718452	142 ( 142 / 0 )	6,4357E-06	687	43276
17	ensimmäisenä	päivänä	11,56674636	135 ( 135 / 0 )	6,1185E-06	2759	1618
18	viime	päivien	10,99543588	123 ( 123 / 0 )	5,5746E-06	295	26304
19	näinä	päivinä	10,90642666	119 ( 119 / 0 )	5,3933E-06	737	252
20	viimeisenä	päivänä	10,88724881	119 ( 119 / 0 )	5,3933E-06	2759	625
21	muutama	päivä	10,88584453	120 ( 120 / 0 )	5,4387E-06	2101	2632
22	kahden	päivän	10,76268622	127 ( 127 / 0 )	5,7559E-06	6492	6470
23	ensimmäisen	päivän	10,42888724	118 ( 118 / 0 )	0,000005348	6492	5340
24	sinä	päivänä	10,17536686	104 ( 104 / 0 )	4,7135E-06	2759	617
25	yhden	päivän	9,843487946	106 ( 106 / 0 )	4,8042E-06	6492	5274
26	puolen	päivän	9,804240516	100 ( 100 / 0 )	4,5322E-06	6492	2218
27	tähän	päivään	9,779015414	96 ( 96 / 0 )	4,3509E-06	465	2939
28	edellisenä	päivänä	9,626555338	93 ( 93 / 0 )	0,000004215	2759	440
29	tuntia	päivässä	9,558845916	92 ( 92 / 0 )	4,1696E-06	977	2371
30	edeltävänä	päivänä	9,48198861	90 ( 90 / 0 )	0,000004079	2759	123
31	yötä	päivää	9,472617827	90 ( 90 / 0 )	0,000004079	2399	414
32	parin	päivän	9,291738245	90 ( 90 / 0 )	0,000004079	6492	2097
33	toisena	päivänä	8,799580144	78 ( 78 / 0 )	3,5351E-06	2759	758
34	syntymäpäivää	päivänä	8,770191666	77 ( 77 / 0 )	3,4898E-06	2759	112
35	julkaistaan	päivänä	8,69655577	76 ( 76 / 0 )	3,4445E-06	2759	494
36	kesäkuun	päivänä	7,949837667	64 ( 64 / 0 )	2,9006E-06	2759	1070
37	jonain	päivänä	7,806562104	61 ( 61 / 0 )	2,7647E-06	2759	77
38	aiemmin	päivällä	7,750349272	62 ( 62 / 0 )	0,00000281	1573	4553
39	kertaa	päivässä	7,689075948	61 ( 61 / 0 )	2,7647E-06	977	7125
41	myöhemmin	päivällä	7,675175336	61 ( 61 / 0 )	2,7647E-06	1573	4933
42	eräänä	päivänä	7,672169816	59 ( 59 / 0 )	0,000002674	2759	184

Kolokačni tabulka "päivä" MI score (-3)							
	Kolokace		MI score	Abs (Nal./Napr.)	Rel.	Celkem nodů	Celkem kolokátů
1	parhaina	päivinään	15,57822388	6 ( 6 / 0 )	2,719E-07	11	82
2	vanhoilla	päivillään	15,42816141	16 ( 16 / 0 )	7,252E-07	17	157
3	ainoan	päivämme	13,94543981	5 ( 5 / 0 )	2,266E-07	7	333
4	viimeisiin	päiviin	13,22228001	5 ( 5 / 0 )	2,266E-07	104	37
5	viimeisistä	päivistä	13,20476548	6 ( 6 / 0 )	2,719E-07	57	82
7	muutamaksi	päiväksi	13,14953596	25 ( 25 / 0 )	1,1331E-06	213	95
8	hyvänä	päivänään	12,78470991	7 ( 7 / 0 )	3,173E-07	12	608
9	aurinkoisina	päivinä	12,69976169	8 ( 8 / 0 )	3,626E-07	737	12
10	lukee	päivämme	12,58672592	5 ( 5 / 0 )	2,266E-07	7	854
11	molempina	päivinä	12,54106395	43 ( 43 / 0 )	1,9489E-06	737	72
12	peräkkäisinä	päivinä	12,51918945	10 ( 10 / 0 )	4,532E-07	737	17
13	näihin	päiviin	12,44431145	45 ( 45 / 0 )	2,0395E-06	104	571
14	onnen	päivät	12,20655641	283 ( 283 / 0 )	1,28262E-05	687	641
15	näinä	päivinä	12,20226203	119 ( 119 / 0 )	5,3933E-06	737	252
16	joinakin	päivinä	12,10029962	11 ( 11 / 0 )	4,985E-07	737	25
17	katsauksen	päivästä	11,98510274	13 ( 13 / 0 )	5,892E-07	265	89
18	barrelia	päivässä	11,97113966	16 ( 16 / 0 )	7,252E-07	977	30
19	pariksi	päiväksi	11,87166305	14 ( 14 / 0 )	6,345E-07	213	129
20	pahoina	päivinä	11,86968669	6 ( 6 / 0 )	2,719E-07	737	16
21	sodoman 120	päivää	11,58202756	13 ( 13 / 0 )	5,892E-07	2399	13
22	kymmeneksi	päiväksi	11,57303504	9 ( 9 / 0 )	4,079E-07	213	102
23	jokaiselle	päivälle	11,51116421	9 ( 9 / 0 )	4,079E-07	58	391
24	neljäksi	päiväksi	11,44326717	10 ( 10 / 0 )	4,532E-07	213	124
25	ikkunanäyttelyssä	päivää	11,34756231	17 ( 17 / 0 )	7,705E-07	2399	20
26	päivästä	päivään	11,32207012	45 ( 45 / 0 )	2,0395E-06	465	278
27	parhaimpina	päivinä	11,21433487	5 ( 5 / 0 )	2,266E-07	737	21
28	iisi	päivä	11,18842206	8 ( 8 / 0 )	3,626E-07	2101	12
29	kaikkina	päivinä	11,09207912	21 ( 21 / 0 )	9,518E-07	737	96
30	meidän	päiviumme	11,04661765	14 ( 14 / 0 )	6,345E-07	14	3477
31	jonain	päivänä	11,04426606	61 ( 61 / 0 )	2,7647E-06	2759	77
32	jonakin	päivänä	11,02884289	58 ( 58 / 0 )	2,6287E-06	2759	74
33	edeltävänä	päivänä	10,92965385	90 ( 90 / 0 )	0,000004079	2759	123
34	viimeisinä	päivinä	10,90885729	21 ( 21 / 0 )	9,518E-07	737	109
35	syntymäpäivää	päivänä	10,83974688	77 ( 77 / 0 )	3,4898E-06	2759	112
36	täksi	päiväksi	10,77341577	11 ( 11 / 0 )	4,985E-07	213	217
37	tieteen	päivillä	10,58110575	15 ( 15 / 0 )	6,798E-07	112	643
38	seitsemänä	päivänä	10,52616613	26 ( 26 / 0 )	1,1784E-06	2759	47
39	joka ikinen	päivä	10,39487293	5 ( 5 / 0 )	2,266E-07	2101	13
41	parhaat	päivänsä	10,33995897	15 ( 15 / 0 )	6,798E-07	53	1606
42	tälle	päivälle	10,26809983	6 ( 6 / 0 )	2,719E-07	58	617

Substantivum *tunti*

Kolokačni tabulka "tunti" t-score (-3/+3)							
	Kolokace			t-score	Abs (Nal./Napr.)	Rel.	Celkem nodů
1	puolen	tunnin		21,59271981	469 ( 469 / 0 )	2,1256E-05	2295
2	puoli	tuntia		18,38701363	342 ( 340 / 2 )	1,55E-05	2371
3	kaksi	tuntia		14,96775311	246 ( 245 / 1 )	1,1149E-05	2371
4		tunnin	välein	14,46457236	210 ( 0 / 210 )	9,5177E-06	2295
5		tuntia	ennen	13,97203382	213 ( 2 / 211 )	9,6536E-06	2371
6		tunnin	jälkeen	12,52677509	189 ( 9 / 180 )	8,5659E-06	2295
7		tunnin	ajan	12,16281012	155 ( 0 / 155 )	7,0249E-06	2295
8	pari	tuntia		11,99674665	150 ( 150 / 0 )	6,7983E-06	2371
9	noin	tunnin		11,59091123	161 ( 157 / 4 )	7,2969E-06	2295
10	kilometriä	tunnissa		11,33458474	129 ( 125 / 4 )	5,8466E-06	576
11	kestää	tuntia		11,30650315	130 ( 130 / 0 )	5,8919E-06	2371
12	naisten	tunti		11,29340652	130 ( 130 / 0 )	5,8919E-06	582
13	noin	tuntia		10,84663858	145 ( 140 / 5 )	6,5717E-06	2371
14	puolitoista	tuntia		10,67567415	115 ( 115 / 0 )	5,2121E-06	2371
15	parin	tunnin		10,1192579	105 ( 103 / 2 )	4,7588E-06	2295
16		tuntia	päivässä	10,03441124	102 ( 8 / 94 )	4,6229E-06	2371
17		tunnin	kuluttua	9,555247452	94 ( 0 / 94 )	4,2603E-06	2295
18		tuntia	viikossa	9,362854207	89 ( 9 / 80 )	4,0337E-06	2371
19	kesti	tuntia		9,319748303	88 ( 86 / 2 )	3,9884E-06	2371
20	24	tunnin		8,89442892	84 ( 81 / 3 )	3,8071E-06	2295
21	yli	tunnin		8,525074327	93 ( 91 / 2 )	4,215E-06	2295
22	24	tuntia		8,482836779	77 ( 76 / 1 )	3,4898E-06	2371
23		tuntia	myöhemmin	8,293007753	75 ( 0 / 75 )	3,3992E-06	2371
24	muutaman	tunnin		8,214483987	71 ( 70 / 1 )	3,2179E-06	2295
25	puolisen	tuntia		8,105003251	66 ( 66 / 0 )	2,9913E-06	2371
26		tunnin	mittainen	8,036957042	* 65 ( 0 / 65 )	2,9459E-06	2295
27	muutama	tunti		7,947942325	64 ( 64 / 0 )	2,9006E-06	582
28		tuntia	minuuttia	7,911355658	66 ( 5 / 61 )	2,9913E-06	2371
29	vajaan	tunnin		7,816080487	62 ( 61 / 1 )	0,00000281	2295
30	useita	tunteja		7,722666537	60 ( 60 / 0 )	2,7193E-06	200
31	runsaan	tunnin		7,615100386	59 ( 59 / 0 )	2,674E-06	2295
32		tuntia	vuorokaudessa	7,3927343	55 ( 3 / 52 )	2,4927E-06	2371
33		tunnin	verran	7,11997883	52 ( 0 / 52 )	2,3568E-06	2295
34		tunnin	aikana	6,816706928	61 ( 0 / 61 )	2,7647E-06	2295
35	kestää	tunnin		6,702272212	47 ( 47 / 0 )	2,1301E-06	2295
36	lähes	tuntia		6,543503016	56 ( 50 / 6 )	2,538E-06	2371
37	lähes	tunnin		6,188190061	51 ( 50 / 1 )	2,3114E-06	2295
38		tunnin	ajomatkan	5,740221359	33 ( 0 / 33 )	1,4956E-06	2295
39		tunnin	mittaisen	5,639427157	32 ( 0 / 32 )	1,4503E-06	2295
40		tuntia	myöhässä	5,620157614	32 ( 0 / 32 )	1,4503E-06	2371

Kolokačni tabulka "tunti" MI score (-3/+3)							
	Kolokace			MI score	Abs (Nal./Napr.)	Rel.	Celkem nodů
1	useiksi	tunneiksi		16,0509116	6 ( 6 / 0 )	2,719E-07	13
2		tunnista	tuntiin	13,02133848	13 ( 0 / 13 )	5,892E-07	46
3	puoleksitoista	tunniksi		12,55083799	7 ( 7 / 0 )	3,173E-07	143
4	muutamaksi	tunniksi		12,40244615	20 ( 20 / 0 )	9,064E-07	143
5	kahteentoista	tuntiin		12,12199469	5 ( 5 / 0 )	2,266E-07	125
6	puoleksi	tunniksi		11,76501623	18 ( 18 / 0 )	8,158E-07	143
7	pariksi	tunniksi		11,72660925	17 ( 17 / 0 )	7,705E-07	143
8	kilometriin	tunnissa		11,54078433	28 ( 27 / 1 )	1,269E-06	576
9	puolessatoista	tunnissa		10,91785398	10 ( 10 / 0 )	4,532E-07	576
10	parilla	tunnilla		10,76585088	7 ( 7 / 0 )	3,173E-07	77
11	reilussa	tunnissa		10,64032	12 ( 12 / 0 )	5,439E-07	576
12	puolessa	tunnissa		10,52675744	58 ( 57 / 1 )	2,6287E-06	576
13	le mansin 24	tunnin		10,45332149	7 ( 7 / 0 )	3,173E-07	2295
14		tunnin	ajomatkan	10,36843259	33 ( 0 / 33 )	1,4956E-06	2295
15		tunnin	työpäivään	10,19850759	11 ( 1 / 10 )	4,985E-07	2295
16	kuuteen	tuntiin		9,937570123	8 ( 8 / 0 )	3,626E-07	125
17	luterilainen	tunti		9,737844389	10 ( 10 / 0 )	4,532E-07	582
18		tunnin	opastuskierros	9,645966565	5 ( 0 / 5 )	2,266E-07	2295
19		tunnin	työpäivästä	9,530489347	6 ( 1 / 5 )	2,719E-07	2295
20	penniä	tunnilta		9,422961976	11 ( 11 / 0 )	4,985E-07	71
21	useiden	tuntien		9,267279113	11 ( 11 / 0 )	4,985E-07	95
22	vajaassa	tunnissa		9,116758045	6 ( 6 / 0 )	2,719E-07	576
23		tunnin	mittaisessa	9,095769482	14 ( 0 / 14 )	6,345E-07	2295
24		tunnin	välein	9,074809863	210 ( 0 / 210 )	9,5177E-06	2295
25	runsaassa	tunnissa		9,021410169	7 ( 7 / 0 )	3,173E-07	576
26	puoleen	tuntiin		9,000763949	14 ( 13 / 1 )	6,345E-07	125
27	muutamia	tunteja		8,982221889	28 ( 28 / 0 )	1,269E-06	200
28		tunnin	mittaiseen	8,935473182	11 ( 0 / 11 )	4,985E-07	2295
29	kilometriä	tunnissa		8,931302999	129 ( 125 / 4 )	5,8466E-06	576
30	muutamassa	tunnissa		8,870439785	39 ( 39 / 0 )	1,7676E-06	576
31	puolisen	tuntia		8,736468579	66 ( 66 / 0 )	2,9913E-06	2371
32		tuntia	kestäneessä	8,65706801	19 ( 0 / 19 )	8,611E-07	2371
33	puolisentoista	tuntia		8,548338982	7 ( 7 / 0 )	3,173E-07	2371
34		tunnin	mittaista	8,432972841	11 ( 0 / 11 )	4,985E-07	2295
35	puolen	tunnin		8,404367027	469 ( 469 / 0 )	2,1256E-05	2295
36		tunnista	toiseen	8,393598251	7 ( 0 / 7 )	3,173E-07	46
37		tunnin	työpäivää	8,382932159	10 ( 1 / 9 )	4,532E-07	2295
38	useita	tunteja		8,376311882	60 ( 60 / 0 )	2,7193E-06	200
39		tunnin	mittainen	8,315187552	65 ( 0 / 65 )	2,9459E-06	2295
40		tuntia	vuorokaudessa	8,303509172	55 ( 3 / 52 )	2,4927E-06	2371

Substantivum *minuutti*

Kolokačni tabulka "minuutti" t-score (-3/+3)							
	Kolokace			t-score	Abs (Nal.Napr.)	Rel.	Celkem nodů
1	kymmenen	minuuttia		17,1688057	306 ( 305 / 1 )	1,3869E-05	2680
2		minuuttia	ennen	16,1813908	282 ( 1 / 281 )	1,2781E-05	2680
3	pientä	minuuttia		14,19048141	203 ( 203 / 0 )	9,2004E-06	2680
4	viisi	minuuttia		14,11028209	207 ( 205 / 2 )	9,3817E-06	2680
5	noin	minuuttia		13,10629309	203 ( 201 / 2 )	9,2004E-06	2680
6	kaksi	minuuttia		12,02263907	169 ( 164 / 5 )	7,6594E-06	2680
7	kolme	minuuttia		11,54263349	151 ( 147 / 4 )	6,8436E-06	2680
8		minuuttia	loppua	11,34841591	130 ( 0 / 130 )	5,8919E-06	2680
9		minuutin	jälkeen	11,06931066	141 ( 8 / 133 )	6,3904E-06	1307
10	viiden	minuutin		10,92221421	122 ( 121 / 1 )	5,5293E-06	1307
11	puoli	minuuttia		10,07968799	106 ( 104 / 2 )	4,8042E-06	2680
12		minuuttia	myöhemmin	10,04639318	108 ( 0 / 108 )	4,8948E-06	2680
13	kahden	minuutin		9,922333825	103 ( 97 / 6 )	4,6682E-06	1307
14	kymmenen	minuutin		9,825702737	102 ( 99 / 3 )	4,6229E-06	1307
15	vain	minuuttia		9,726925849	139 ( 137 / 2 )	6,2998E-06	2680
16	20	minuuttia		9,569094201	114 ( 111 / 3 )	5,1667E-06	2680
17	yli	minuuttia		9,017633477	105 ( 79 / 26 )	4,7588E-06	2680
18		minuutin	aikana	8,90975574	88 ( 4 / 84 )	3,9884E-06	1307
19	pari	minuuttia		8,843058913	85 ( 85 / 0 )	3,8524E-06	2680
20	15	minuuttia		8,822601969	100 ( 98 / 2 )	4,5322E-06	2680
21	asteessa	minuuttia		8,413276247	71 ( 71 / 0 )	3,2179E-06	2680
22	puolitoista	minuuttia		8,114048275	67 ( 67 / 0 )	3,0366E-06	2680
23	kestää	minuuttia		8,097307413	68 ( 67 / 1 )	3,0819E-06	2680
24		minuutin	kuluttua	7,966298624	65 ( 0 / 65 )	2,9459E-06	1307
25	tuntia	minuuttia		7,911355658	66 ( 61 / 5 )	2,9913E-06	2680
26	muutaman	minuutin		7,873038861	64 ( 64 / 0 )	2,9006E-06	1307
27	neljä	minuuttia		7,795894456	70 ( 68 / 2 )	3,1726E-06	2680
28	kolmen	minuutin		7,767835414	64 ( 62 / 2 )	2,9006E-06	1307
29		minuutin	pelin	7,607815922	59 ( 1 / 58 )	2,674E-06	1307
30	lähes	minuuttia		7,548431491	72 ( 69 / 3 )	3,2632E-06	2680
31		minuuttia	sydämellä	7,336081395	54 ( 0 / 54 )	2,4474E-06	2680
32	kesti	minuuttia		7,328846691	55 ( 55 / 0 )	2,4927E-06	2680
33		minuutin	välein	7,039711372	50 ( 0 / 50 )	2,2661E-06	1307
34	45	minuuttia		7,008131061	53 ( 52 / 1 )	2,4021E-06	2680
35		minuutin	ajan	6,928792457	52 ( 2 / 50 )	2,3568E-06	1307
36	puolen	minuutin		6,88739138	49 ( 47 / 2 )	2,2208E-06	1307
37		minuutin	sekuntia	6,702264713	46 ( 3 / 43 )	2,0848E-06	1307
38	jo	minuutilla		6,70204722	55 ( 49 / 6 )	2,4927E-06	545
39	muutamassa	minuutissa		6,623812636	44 ( 44 / 0 )	1,9942E-06	433

Kolokačni tabulka "minuutti" MI score (-3/+3)							
	Kolokace			MI score	Abs (Nal./Napr.)	Rel.	Celkem nodů
1	minuutista	minuuttiin		12,81306552	5 ( 5 / 0 )	2,266E-07	73
2		minuutista	tuntiin	12,0371058	5 ( 0 / 5 )	2,266E-07	35
3	muutamaksi	minuutiksi		11,9673709	6 ( 6 / 0 )	2,719E-07	58
4	viideksi	minuutiksi		11,90786989	8 ( 8 / 0 )	3,626E-07	58
5	viidessätoista	minuutissa		11,77191387	7 ( 7 / 0 )	3,173E-07	433
6	kymmeneksi	minuutiksi		11,60176676	5 ( 5 / 0 )	2,266E-07	58
7		minuutti	minuutilta	11,5773102	7 ( 0 / 7 )	3,173E-07	156
8	25+25	minuutin		11,45820158	5 ( 5 / 0 )	2,266E-07	1307
9	runsaassa	minuutissa		11,01807446	21 ( 21 / 0 )	9,518E-07	433
10	vajaassa	minuutissa		10,94349733	16 ( 16 / 0 )	7,252E-07	433
11		minuutin	aikasakko	10,78012967	5 ( 0 / 5 )	2,266E-07	1307
12	30+30	minuutin		10,72123598	6 ( 6 / 0 )	2,719E-07	1307
13	reilussa	minuutissa		10,63698429	9 ( 9 / 0 )	4,079E-07	433
14	penniä	minuutilta		10,52831498	18 ( 18 / 0 )	8,158E-07	54
15	kymmenestä	minuutista		10,48325383	5 ( 5 / 0 )	2,266E-07	35
16		minuutin	hälytyksestä	10,45820158	6 ( 0 / 6 )	2,719E-07	1307
17		minuuttia	päättösvihellystä	10,29669684	11 ( 0 / 11 )	4,985E-07	2680
18	viimeiseen	minuuttiin		9,956100152	7 ( 7 / 0 )	3,173E-07	73
19	viiteen	minuuttiin		9,892499989	7 ( 7 / 0 )	3,173E-07	73
20	johtomaalin	minuutilla		9,862151588	8 ( 6 / 2 )	3,626E-07	545
21	muutamia	minuutteja		9,855249033	10 ( 10 / 0 )	4,532E-07	39
22		minuutin	kävelymatkan	9,814345387	8 ( 0 / 8 )	3,626E-07	1307
23	uuninlämmössä	minuuttia		9,685262123	9 ( 9 / 0 )	4,079E-07	2680
24	maalinteon	minuutilla		9,684508673	5 ( 5 / 0 )	2,266E-07	545
25	viimeiset	minuutit		9,682787978	9 ( 9 / 0 )	4,079E-07	31
26		minuutin	alitus	9,650846655	6 ( 0 / 6 )	2,719E-07	1307
27	kymmenessä	minuutissa		9,582536505	26 ( 26 / 0 )	1,1784E-06	433
28	kuudennella	minuutilla		9,577174629	6 ( 6 / 0 )	2,719E-07	545
29	muutamassa	minuutissa		9,456170971	44 ( 44 / 0 )	1,9942E-06	433
30		minuutin	jäähyn	9,399307888	12 ( 0 / 12 )	5,439E-07	1307
31	asteessa	minuuttia		9,352806317	71 ( 71 / 0 )	3,2179E-06	2680
32	viidennellä	minuutilla		9,350898774	9 ( 9 / 0 )	4,079E-07	545
33	maalintekoon	minuutilla		9,287173176	5 ( 3 / 2 )	2,266E-07	545
34	viidessä	minuutissa		9,261384261	31 ( 29 / 2 )	1,405E-06	433
35		minuutin	mittaisia	9,248748211	16 ( 0 / 16 )	7,252E-07	1307
36		minuuttia	sydämellä	9,211330935	54 ( 0 / 54 )	2,4474E-06	2680



Substantivum *sekunti*

Kolokační tabulka "sekunti" t-score (-3/+3)							
	Kolokace			t-score	Abs (Nal./Napr.)	Rel.	Celkem nodů
1		sekuntia	ennen	11,58122753	138 ( 0 / 138 )	6,2545E-06	1528
2	metriä	sekunnissa		11,52536641	133 ( 133 / 0 )	6,0278E-06	342
3		sekuntia	jäljessä	9,265437055	86 ( 0 / 86 )	3,8977E-06	1528
4		sekunnin	erolla	7,74400602	60 ( 0 / 60 )	2,7193E-06	658
5	kaksi	sekuntia		7,501073081	61 ( 60 / 1 )	2,7647E-06	1528
6	kolme	sekuntia		6,969638729	52 ( 51 / 1 )	2,3568E-06	1528
7		sekunnin	päähän	6,917156968	48 ( 0 / 48 )	2,1755E-06	658
8	kymmenen	sekuntia		6,846049866	49 ( 49 / 0 )	2,2208E-06	1528
9	neljä	sekuntia		6,099689969	39 ( 39 / 0 )	1,7676E-06	1528
10	viisi	sekuntia		5,873642445	36 ( 36 / 0 )	1,6316E-06	1528
11	15	sekuntia		5,62705792	36 ( 36 / 0 )	1,6316E-06	1528
12	muutaman	sekunnin		5,53715228	31 ( 31 / 0 )	1,405E-06	658
13		sekuntia	myöhemmin	5,536076806	32 ( 0 / 32 )	1,4503E-06	1528
14		sekunnin	päässä	5,363415832	29 ( 0 / 29 )	1,3143E-06	658
15	muutamassa	sekunnissa		5,288389245	28 ( 28 / 0 )	1,269E-06	342
16	20	sekuntia		5,259403086	32 ( 32 / 0 )	1,4503E-06	1528
17	kymmenen	sekunnin		5,008006951	26 ( 26 / 0 )	1,1784E-06	658
18		sekuntia	nopeampi	4,791126548	23 ( 0 / 23 )	1,0424E-06	1528
19	11	sekuntia		4,769399521	26 ( 26 / 0 )	1,1784E-06	1528
20	14	sekuntia		4,663599639	26 ( 26 / 0 )	1,1784E-06	1528
21		sekuntia	perässä	4,661272688	22 ( 0 / 22 )	9,971E-07	1528
22	kahdeksan	sekuntia		4,596396836	22 ( 20 / 2 )	9,971E-07	1528
23	puoli	sekuntia		4,377646742	20 ( 20 / 0 )	9,064E-07	1528
24	kuusi	sekuntia		4,249848526	19 ( 19 / 0 )	8,611E-07	1528
25	neljän	sekunnin		4,204783608	18 ( 18 / 0 )	8,158E-07	658
26		sekunnin	sadasosaa	4,122253732 *	17 ( 0 / 17 )	7,705E-07	658

Kolokačnī tabulka "sekunti" MI score (-3/+3)							
	Kolokace			MI score	Abs (Nal./Napr.)	Rel.	Celkem nodû
1	viimeisillā	sekunneilla		15,62042044	6 ( 6 / 0 )	2,719E-07	18
2	megabittiā	sekunnissa		14,49192816	5 ( 5 / 0 )	2,266E-07	342
3	bittiā	sekunnissa		14,12935808	5 ( 5 / 0 )	2,266E-07	342
4		sekunnin	kymmenyksellā	12,89576021	5 ( 0 / 5 )	2,266E-07	658
5		sekunnin	alitusta	12,84061865	7 ( 0 / 7 )	3,173E-07	658
6		sekunnin	sadasosan	12,44830123	10 ( 0 / 10 )	4,532E-07	658
7		sekunnissa	puhaltanut	12,32527829	7 ( 0 / 7 )	3,173E-07	342
8		sekunnin	sadasosaa	12,23808352	17 ( 0 / 17 )	7,705E-07	658
9	muutamia	sekunteja		12,01136823	8 ( 8 / 0 )	3,626E-07	21
10		sekunnin	erolla	11,94015433	60 ( 0 / 60 )	2,7193E-06	658
11		sekunnin	murto-osan	11,61822623	6 ( 0 / 6 )	2,719E-07	658
12		sekunnin	kymmenystä	11,33282401	8 ( 0 / 8 )	3,626E-07	658
13		sekunnin	etumatkan	11,1126982	7 ( 0 / 7 )	3,173E-07	658
14		sekunnin	kymmenyksen	10,9628744	10 ( 0 / 10 )	4,532E-07	658
15	metriin	sekunnissa		10,93921986	7 ( 7 / 0 )	3,173E-07	342
16	muutamassa	sekunnissa		10,72942748	28 ( 28 / 0 )	1,269E-06	342
17	metriā	sekunnissa		10,63924697	133 ( 133 / 0 )	6,0278E-06	342
18	viimeisellā	sekunnilla		10,58503757	11 ( 11 / 0 )	4,985E-07	133
19		sekuntia	jäljessä	10,14338109	86 ( 0 / 86 )	3,8977E-06	1528
20		sekuntia	nopeampi	9,992612477	23 ( 0 / 23 )	1,0424E-06	1528
21		sekunnin	etumatka	9,735583181	6 ( 0 / 6 )	2,719E-07	658
22	muutama	sekunti		9,317056696	7 ( 7 / 0 )	3,173E-07	46
23		sekunnin	päähän	9,291796743	48 ( 0 / 48 )	2,1755E-06	658
24		sekunnin	eron	8,488943214	9 ( 0 / 9 )	4,079E-07	658
25		sekunnin	välein	8,069789606	10 ( 0 / 10 )	4,532E-07	658
26		sekunnin	päässä	7,951665511	29 ( 0 / 29 )	1,3143E-06	658
27	viime	sekunneilla		7,864216013	10 ( 10 / 0 )	4,532E-07	18
28		sekuntia	huonempi	7,832885568	6 ( 0 / 6 )	2,719E-07	1528
29		sekunnin	johdossa	7,679626775	9 ( 0 / 9 )	4,079E-07	658
30	muutaman	sekunnin		7,506669839	31 ( 31 / 0 )	1,405E-06	658
31	kymmenisen	sekuntia		7,455834902	8 ( 8 / 0 )	3,626E-07	1528
32		sekuntia	perässä	7,33030402	22 ( 0 / 22 )	9,971E-07	1528
33	kuuden	sekunnin		7,195514516	13 ( 13 / 0 )	5,892E-07	658
34		sekunnin	verran	7,16289901	9 ( 0 / 9 )	4,079E-07	658
35	kilometriā	sekunnissa		6,579038825	5 ( 5 / 0 )	2,266E-07	342

Substantivum *kuu*

Kolokačni tabulka "kuu" t-score (-3/+3)							
	Kolokace			t-score	Abs (Nal./Napr.)	Rel.	Celkem nodů
1	markkaa	kuussa		15,13126923	240 ( 223 / 17 )	1,0877E-05	955
2	viime	kuussa		14,23221431	216 ( 216 / 0 )	9,7896E-06	955
3	tässä	kuussa		13,1933338	177 ( 177 / 0 )	8,022E-06	955
4	tämän	kuun		12,30247374	154 ( 154 / 0 )	6,9796E-06	565
5	ensi	kuussa		10,98237485	128 ( 125 / 3 )	5,8012E-06	955
6		kuun	lopussa	9,225209194	86 ( 0 / 86 )	3,8977E-06	565
7	kerran	kuussa		8,308211635	72 ( 70 / 2 )	3,2632E-06	955
8		kuun	alussa	7,757138433	62 ( 1 / 61 )	0,00000281	565
9		kuun	lopulla	7,232950292	53 ( 0 / 53 )	2,4021E-06	565
10	ensi	kuun		7,117135177	55 ( 53 / 2 )	2,4927E-06	565
11		kuun	päivänä	6,335056652	41 ( 2 / 39 )	1,8582E-06	565
12	aiemmin	kuussa		6,137607613	40 ( 40 / 0 )	1,8129E-06	955
13	viime	kuun		5,857144071	42 ( 42 / 0 )	1,9035E-06	565
14		kuun	loppuun	5,382781188	30 ( 2 / 28 )	1,3597E-06	565
15	noin	kuussa		4,836967424	34 ( 28 / 6 )	1,541E-06	955
16		kuun	alusta	4,810446948	24 ( 1 / 23 )	1,0877E-06	565
17	kuluvan	kuun		4,77904683	23 ( 23 / 0 )	1,0424E-06	565
18		kuun	vaihteessa	4,772543394	23 ( 0 / 23 )	1,0424E-06	565
19		kuun	aikana	4,724364584	26 ( 0 / 26 )	1,1784E-06	565
20		kuun	puolivälissä	4,547541648	21 ( 0 / 21 )	9,518E-07	565

Kolokační tabulka "kuu" MI score (-3/+3)							
	Kolokace			MI score	Abs (Nat./Napr.)	Rel.	Celkem nodů
1		kuulla	raskaana	12,01810545	5 ( 1 / 4 )	2,266E-07	31
2		kuuhun	keravan	10,66424584	19 ( 0 / 19 )	8,611E-07	46
3		kuussa	lapselta	8,91088808	6 ( 0 / 6 )	2,719E-07	955
4	kulovan	kuun		8,1582769	23 ( 23 / 0 )	1,0424E-06	565
5		kuun	vaihteessa	7,685888348	23 ( 0 / 23 )	1,0424E-06	565
6		kuussa	presidentinvaalit	7,675671618	6 ( 1 / 5 )	2,719E-07	955
7		kuussa	täyttävä	7,672483341	8 ( 0 / 8 )	3,626E-07	955
8		kuun	lopussa	7,581168586	86 ( 0 / 86 )	3,8977E-06	565
9	tätä	kuuta		7,580705145	10 ( 10 / 0 )	4,532E-07	58
10		kuun	lopulla	7,269999284	53 ( 0 / 53 )	2,4021E-06	565
11	4500	kuussa		7,134784092	5 ( 5 / 0 )	2,266E-07	955
12		kuun	puolivälissä	7,031168141	21 ( 0 / 21 )	9,518E-07	565
13	tässä	kuussa		6,907072434	177 ( 177 / 0 )	8,022E-06	955
14	tämän	kuun		6,854440029	154 ( 154 / 0 )	6,9796E-06	565
15	päivänä	kuuta		6,804041667	5 ( 5 / 0 )	2,266E-07	58
16		kuun	päivänä	6,555539515	41 ( 2 / 39 )	1,8582E-06	565
17		kuun	vaihteen	6,52246049	5 ( 0 / 5 )	2,266E-07	565
18		kuun	loppupuolella	6,382735727	5 ( 0 / 5 )	2,266E-07	565
19	1200	kuussa		6,291978247	7 ( 7 / 0 )	3,173E-07	955
20		kuun	viimeisenä	6,187780488	7 ( 0 / 7 )	3,173E-07	565
21		kuun	alkupuolella	6,113549094	5 ( 0 / 5 )	2,266E-07	565
22		kuun	alussa	6,073993968	62 ( 1 / 61 )	0,00000281	565
23		kuun	loppuun	5,857780535	30 ( 2 / 28 )	1,3597E-06	565
24	virallisesti	kuun		5,842497671	7 ( 7 / 0 )	3,173E-07	565
25	syksyn	kuu		5,820096709	6 ( 6 / 0 )	2,719E-07	276
26		kuun	alusta	5,790087033	24 ( 1 / 23 )	1,0877E-06	565
27	kerran	kuussa		5,582462473	72 ( 70 / 2 )	3,2632E-06	955
28	päättyy	kuun		5,580675104	11 ( 10 / 1 )	4,985E-07	565
29	tuhat	kuussa		5,443909652	13 ( 11 / 2 )	5,892E-07	955
30	markkaa	kuussa		5,424388218	240 ( 223 / 17 )	1,0877E-05	955
31	kokoontuu	kuussa		5,353232925	5 ( 5 / 0 )	2,266E-07	955
32	voimaan	kuun		5,214867312	8 ( 8 / 0 )	3,626E-07	565
33	viimeistään	kuun		5,164312208	6 ( 6 / 0 )	2,719E-07	565
34	ensi	kuussa		5,093504652	128 ( 125 / 3 )	5,8012E-06	955
35	4000	kuussa		5,093264822	5 ( 5 / 0 )	2,266E-07	955
36	aiemmin	kuussa		5,080214431	40 ( 40 / 0 )	1,8129E-06	955

## **Příloha č. 2: Kolokační tabulky substantiv času v češtině**

Kolokační tabulky substantiv času v češtině byly vygenerovány z korpusu SYN2000. Při jejich sestavování byly přednastaveny tyto parametry: kontext výskytu kolokátů byl omezen na -5/+5 slov od nodu (5 slov nalevo a napravo od KWIC)., minimální výskyt v korpusu byl nastaven na hodnotu 5.

Tabulky jsou sestaveny podle klesajících hodnot míry asociace (t-score a MI score). Kromě hodnot míry asociace uvádějí také relativní a absolutní frekvenci daných kolokací.

Kolokační tabulky substantiva čas (t-score / MI score)

Word	T-score	Rel. f [%]	Abs. f	word	MI-score	Rel. f [%]	Abs.f
reálném	18.2	49.78	332	mimoherního	11.51	150	9
postupem	14.34	67.76	206	šetříme	10.93	100	13
nastaveném	11.78	91.45	139	nemarní	10.93	100	5
nezbyl	10.48	61.45	110	dočkej	10.93	100	7
oddechový	10.19	85.25	104	hlodů	10.93	100	37
nadešel	9.891	59.76	98	Julliette	10.93	100	5
vysílacího	9.79	64.43	96	mar nil	10.93	100	6
plynul	8.879	49.69	79	drahný	10.93	100	6
vysílacím	8.654	70.75	75	rekordním	10.93	100	5
Archivy	8.538	73	73	oddechový	10.8	91.67	11
nejkratším	8.476	45.86	72	nastaveném	10.8	91.45	139
strávený	7.408	46.22	55	stopuje	10.7	85.71	6
blýská	7.133	44.35	51	oddechový	10.7	85.25	104
Rozmary	6.923	66.67	48	dozrál	10.66	83.33	10
Nadešel	6.628	62.86	44	greenwichského	10.64	81.82	9
plynutí	6.474	48.84	42	marníte	10.6	80	4
Hlodů	6.08	100	37	odměřovaly	10.6	80	4
zašlých	5.738	47.83	33	provolaného	10.6	80	4
nastaveného	5.563	57.41	31	mythický	10.6	80	4
azimut	5.473	66.67	30	nemarnil	10.6	80	4
ubíhal	5.38	53.7	29	oddechovými	10.6	80	4
nezbýval	5.289	93.33	28	nevybyl	10.6	80	4
Rumburská	4.997	89.29	25	oddechového	10.56	77.78	7
marnit	4.687	75.86	22	ztraceného	10.55	76.92	10
blýskat	4.685	46.81	22	marnit	10.53	75.86	22
blýskání	4.239	56.25	18	proměškany	10.51	75	6
stráveného	3.869	48.39	15	greenwichský	10.51	75	6
vyměřený	3.738	60.87	14	oddychový	10.51	75	9
šetříme	3.604	100	13	uplývajícího	10.51	75	6
Oddechový	3.315	91.67	11	archivy	10.47	73	73
oddechovém	3.314	61.11	11	trávený	10.47	72.73	8
kilometrovém	3.313	44	11	marnění	10.44	71.43	5
dozrál	3.16	83.33	10	nenadešel	10.44	71.43	10
ztraceného	3.16	76.92	10	vysílacím	10.43	70.75	75
nenadešel	3.16	71.43	10	nazrál	10.39	69.23	9
marní	3.16	66.67	10	postupem	10.36	67.76	206
promarněný	3.16	58.82	10	neztrácel	10.34	66.67	4
mimoherního	2.999	150	9	tráviš	10.34	66.67	4
greenwichského	2.998	81.82	9	ubíhajícího	10.34	66.67	4
oddychový	2.998	75	9	uplývající	10.34	66.67	6
nazrál	2.998	69.23	9	zabíjet	10.34	66.67	4
trávený	2.826	72.73	8	rozmary	10.34	66.67	48
kvapí	2.826	57.14	8	azimut	10.34	66.67	30
uplývání	2.826	50	8	marní	10.34	66.67	10
nepřejícím	2.826	50	8	vysílacího	10.29	64.43	96
vodnes	2.825	44.44	8	nadešel	10.26	62.86	44
dočkej	2.644	100	7				
strávený	7.408	46.22	55				

Kolokační tabulky substantiva *rok* (t-score/MI score)

Word	T-score	Rel. f [%]	Abs. f	word	MI-score	Rel. f [%]	Abs. f
loňském	63.17	78.43	4025	supervolebním	8.478	120	6
šedesátých	56.17	97.75	3177	Kristove	8.438	116.7	14
osmdesátých	50.21	97.35	2539	LOŇSKÉHO	8.438	116.7	7
padesátých	48.05	97.12	2325	Šedesátá	8.243	102	52
sedmdesátých	47.87	98.38	2307	letoucí	8.215	100	20
devadesátých	43.1	98.37	1870	1935	8.215	100	7
třicátých	35.58	98.08	1275	Devadesátá	8.215	100	38
dvacátých	31.16	97.12	978	ZDVOJNÁSOBIT	8.215	100	5
2005	23.82	84.87	572	22347.7	8.215	100	8
2010	22.45	86.39	508	Theakston	8.215	100	5
čtyřicátých	21.74	94.44	476	Vapas	8.215	100	10
báječná	13.43	78.45	182	sedmdesátými	8.215	100	10
šedesátá	12.24	92.07	151	1875	8.215	100	5
šakali	11.45	100	132	učednická	8.215	100	16
2015	10.77	78	117	babího	8.215	100	49
padesátá	10.63	83.82	114	27347.7	8.215	100	6
půldruhým	9.909	81.15	99	Šakali	8.215	100	132
devadesátém	9.55	77.97	92	32355.6	8.215	100	6
fiskálním	9.449	84.91	90	osmdesátými	8.215	100	12
osmdesátá	8.573	97.37	74	Osmdesátá	8.215	100	31
sedmdesátá	8.396	94.67	71	Rozmarného	8.215	100	11
devadesátá	8.216	90.67	68	UPLYNULÝ	8.215	100	7
Šedesátá	7.187	102	52	odsloužená	8.215	100	5
babího	6.976	100	49	Talentem	8.215	100	6
devadesátá	6.144	100	38	ÉTA	8.215	100	5
padesátá	5.721	82.5	33	POTĚ	8.215	100	5
šakalích	5.638	100	32	Liturgický	8.215	100	13
osmdesátá	5.549	100	31	Šakalích	8.215	100	32
Ssedmdesátá	5.366	93.55	29	sedmdesátých	8.192	98.38	2307
ukazatele	5.177	93.1	27	devadesátých	8.192	98.37	1870
osmačtyřicátého	4.674	95.65	22	třicátých	8.187	98.08	1275
výsluhu	4.565	87.5	21	šedesátých	8.182	97.75	3177
přestupný	4.564	84	21	osmdesátá	8.177	97.37	74
letoucí	4.457	100	20	osmdesátých	8.177	97.35	2539
těžního	4.341	82.61	19	dvacátých	8.173	97.12	978
učednická	3.987	100	16	padesátých	8.173	97.12	2325
báječných	3.985	88.89	16	osmačtyřicátého	8.151	95.65	22
Kristove	3.731	116.7	14	sedmdesátá	8.136	94.67	71
foťbalistou	3.726	82.35	14	čtyřicátých	8.133	94.44	476
1986	3.725	77.78	14	Sedmdesátá	8.119	93.55	29
liturgický	3.593	100	13	UKAZATELE	8.112	93.1	27
odsloužených	3.592	86.67	13	šedesátá	8.096	92.07	151
osmdesátými	3.452	100	12	1969	8.09	91.67	11
dačických	3.451	85.71	12	šedesátými	8.09	91.67	11
dvaadevadesátého	3.45	80	12	padesátými	8.09	91.67	11
rozumného	3.305	100	11	zastiňovací	8.078	90.91	10
1969	3.304	91.67	11	devadesátá	8.074	90.67	68
šedesátými	3.304	91.67	11	1950	8.045	88.89	8

Kolokační tabulky substantiva *měsíc* (t-score / MI score)

word	T-score	Rel. f [%]	Abs. f	word	MI-score	Rel. f [%]	Abs. f
letních	20.5	33.1	421	SPORTOVEC	11.6	115.3	68
zimních	14.98	22.82	225	630Kč	11.4	100	6
kalendářní	13.18	35.01	174	losujeme	11.35	96.55	28
jarních	10.8	19.09	117	OBLEČENÍ	11.05	78.57	33
kalendářního	10.52	28.53	111	tišři	10.91	71.43	5
předchozímu	10.18	18.06	104	kalendářnímu	10.89	70.37	19
necelý	9.831	20.68	97	Hořký	10.82	67.07	55
SPORTOVEC	8.244	115.3	68	Průkazka	10.81	66.67	4
nepodmíněně	7.987	23.44	64	postního	10.73	63.16	12
kalendářním	7.669	22.87	59	ramadán	10.66	60	27
Hořký	7.412	67.07	55	Pomijemu	10.66	60	3
započatý	6.319	43.01	40	neusnese	10.66	60	3
kalendářních	6.075	29.13	37	dorůstal	10.66	60	3
OBLEČENÍ	5.742	78.57	33	BI18	10.66	60	3
ramadánu	5.38	40.28	29	Neuběhl	10.4	50	3
losujeme	5.289	96.55	28	započatý	10.18	43.01	40
necelým	5.281	19.44	28	nejstudenějším	10.17	42.86	3
prázdninových	5.281	19.05	28	ramadánu	10.08	40.28	29
ramadán	5.193	60	27	čtyřikrát	9.981	37.5	3
Necelé	4.465	22.22	20	PRAZDNIN	9.981	37.5	6
kalendářnímu	4.357	70.37	19	odsloužili	9.937	36.36	4
postní	4.353	26.03	19	kalendářní	9.882	35.01	174
smolaře	4.238	33.96	18	lunárních	9.66	1.73	30
Necelý	3.994	22.86	16	trvajícimi	9.631	2.233	29.41
postního	3.462	63.16	12	kalendářních	9.617	6.075	29.13
Chladný	3.31	18.33	11	kalendářního	9.587	10.52	28.53
Uplynuly	2.824	23.53	8	postní	9.455	4.353	26.03

Kolokační tabulky substantiva *týden* (t-score / MI score)

word	T-score	Rel. f [%]	Abs. f	word	MI-score	Rel. f [%]	Abs. f
minulý	68.16	65.48	4652	minulý	12.13	183.1	130
minulého	42.5	25.43	1812	Stonožkový	11.26	100	12
Minulý	32.75	68.02	1074	ÚHLEM	11.22	97.56	40
uplynulém	16.04	24.04	258	ZORNÝM	11.22	97.56	40
MINULÝ	11.4	183.1	130	MOUDROST	11.14	92.5	37
uplynulého	11.16	24.18	125	kvizy	11.14	92.31	12
Prostějovského	10.53	76.55	111	UPLYNULEHO	11.12	90.91	10
podmíněně	9.524	25.28	91	PRANOSTIKY	11.1	90	9
sportky	8.824	48.75	78	BURZOVNÍ	11.09	88.89	40
předminulý	8.18	59.29	67	Prostějovském	11.01	84.38	27
Výrok	8.051	30.09	65	sázkovém	10.97	81.82	9
Sportky	7.537	23.65	57	pašijový	10.93	80	4
zveřejníme	7.473	29.95	56	Prostějovského	10.87	76.55	111
tažena	7.409	41.35	55	Minulý	10.7	68.02	1074
kuponem	7.137	60.71	51	minulý	10.65	65.48	4652
Perly	7.063	36.23	50	Aktualita	10.64	65	13
kalendářium	6.774	32.17	46	kuponem	10.54	60.71	51
ÚHLEM	6.322	97.56	40	předminulý	10.5	59.29	67
ZORNÝM	6.322	97.56	40	Předminulý	10.43	56.52	13
BURZOVNÍ	6.322	88.89	40	Motomagazín	10.43	56.25	9
Distributor	6.319	46.51	40	Královnička	10.41	55.56	5
MOUDROST	6.08	92.5	37	štvánický	10.41	55.56	5



Kolokační tabulky substantiva *den* (t-score / MI score)

word	T-score	Rel. f [%]	Abs. f	word	MI-score	Rel. f [%]	Abs. f
všední	25.57	57.39	656	schůzovních	10.32	116.7	7
Jednoho	25.47	58.07	651	Štědrým	10.26	112.1	120
ŠTASTNÝ	22.72	96.46	517	VŠEDNIHO	10.1	100	27
Štědrý	20.85	73.52	436	Sudetoněmeckých	10.1	100	16
všedního	15.47	62.18	240	Děnisoviče	10.1	100	14
Ozvěny	13.98	68.53	196	Brej	10.1	100	7
Štědrým	10.95	112.1	120	Heslovitý	10.1	100	6
REKORDY	9.633	80.87	93	nechval	10.1	100	5
Následujícího	9.154	73.68	84	Adventistů	10.1	100	5
Příštího	9.154	71.79	84	Stloukám	10.1	100	5
kalendářních	8.875	62.2	79	rozbřeskl	10.1	100	5
Štědrého	7.474	71.79	56	OTEVŘENÝCH	10.06	97.67	42
ošetřovací	6.774	70.77	46	ŠTASTNÝ	10.05	96.46	517
OTEVŘENÝCH	6.475	97.67	42	Hracimi	9.961	90.91	10
DVEŘÍ	6.473	75	42	Betonářské	9.946	90	9
slunečný	6.074	60.66	37	pošmourný	9.928	88.89	8
rozděleně	5.907	57.38	35	republiko	9.928	88.89	8
slunečných	5.56	62	31	Inženýrský	9.906	87.5	14
Prostějovských	5.379	78.38	29	Perný	9.887	86.36	19
VŠEDNIHO	5.191	100	27	ošetřovacího	9.887	86.36	19
parných	5.093	83.87	26	brej	9.876	85.71	6
Květinový	4.684	64.71	22	trifidů	9.876	85.71	6
Perný	4.354	86.36	19	parných	9.845	83.87	26
ošetřovacího	4.354	86.36	19	Danske	9.835	83.33	5
VÝROK	4.118	80.95	17	usměvavýho	9.835	83.33	5
Sudetoněmeckých	3.996	100	16	Hracím	9.835	83.33	5
natačecích	3.995	76.19	16	Saló	9.818	82.35	14
pisčítým	3.995	69.57	16	VÝROK	9.794	80.95	17
kamenitým	3.867	60	15	REKORDY	9.792	80.87	93

Kolokační tabulky substantiva *hodina* (t-score / MI score)

word	T-score	Rel. f [%]	Abs. f	word	MI-score	Rel. f [%]	Abs. f
ranních	21.55	73.81	465	přesčasových	10.61	100	39
odpoledních	19.06	78.11	364	Odbavuje	10.61	100	5
dopoledních	14.51	65.94	211	ODPOLEDNÍCH	10.61	100	7
časných	12.95	78.87	168	nalétaných	10.5	92.31	24
vyučovacích	8.477	67.29	72	KLET	10.49	91.67	11
podvečerních	6.776	74.19	46	odučených	10.46	90	9
přesčasových	6.241	100	39	jedenadvacátou	10.37	84.62	11
nalétaných	4.896	92.31	24	ordinačních	10.35	83.33	5
osmnáctou	4.895	70.59	24	brigádnických	10.29	80	8
ordinační	4.894	64.86	24	neomluvených	10.29	80	12
devatenáctou	4.468	71.43	20	suplované	10.29	80	4
neomluvených	3.461	80	12	ROZVRH	10.29	80	4
jedenadvacátou	3.314	84.62	11	časných	10.27	78.87	168
nalétáno	3.314	73.33	11	odpoledních	10.26	78.11	364
dvaadvacátou	3.313	61.11	11	předplatitelská	10.25	77.78	7
návštěvních	3.313	57.89	11	Tanečních	10.25	77.78	7
odučených	2.998	90	9	Půldruhé	10.2	75	9

Kolokační tabulky substantiva *minuta* (t-score / MI score)

word	T-score	Rel. f [%]	Abs. f	word	MI-score	Rel. f [%]	Abs. f
700	27.65	23.61	766	předehtříváme	11.72	90.91	10
vaříme	22.94	41.53	527	odstát	11.69	89.29	375
Úprava	19.64	61.27	386	odpočítáváme	11.69	88.89	16
odstát	19.36	89.29	375	Zhotovení	11.2	63.38	45
Necháme	15.45	40.44	239	louhovat	11.18	62.5	20
MAXIMUM	15.09	29.77	228	Úprava	11.15	61.27	386
18.50	15.02	26.71	226	Rexeso	11.11	59.65	34
Množství	14.23	26.09	203	pražíme	11.1	59.09	13
Párty	13.29	37.74	177	KALORIE	11.02	56	28
Nastavení	10.29	49.3	106	bobtnat	11.01	55.56	10
Pečeme	8.477	29.39	72	zahříváme	11	55.46	66
zahříváme	8.12	55.46	66	Úvodních	10.96	53.85	7
přibývajícimi	7.674	31.05	59	Neuběhlo	10.86	50	3
grilování	7.608	25.44	58	Nastavení	10.83	49.3	106
Vaříme	7.203	22.81	52	devadesáté	10.76	46.67	21
Zhotovení	6.705	63.38	45	vyluhovat	10.74	46.15	12
KALORIE	5.289	56	28	důste	10.72	45.45	5
Přestávka	5.285	22.4	28	vdechů	10.72	45.45	5
Dusíme	5.192	33.75	27	nastavovalo	10.63	42.86	3
opékáme	4.995	25.51	25	tepů	10.63	42.86	18
odstátí	4.791	28.05	23	Druhých	10.63	42.86	3
grilujeme	4.686	26.19	22	&dollar;	10.59	41.62	407
devadesáté	4.58	46.67	21	vaříme	10.59	41.53	527
louhovat	4.47	62.5	20	zapékáme	10.56	40.82	20
zapékáme	4.469	40.82	20	Necháme	10.55	40.44	239
tepů	4.24	42.86	18	smažte	10.53	40	4
rozmrazování	4.118	23.29	17	Párty	10.45	37.74	177
odpočítáváme	3.999	88.89	16	lipovské	10.44	37.5	3
nastaveného	3.996	29.63	16	VAJEČNÉ	10.44	37.5	6

## Seznam zkratek:

ČNK – Český národní korpus

WSE – Word Sketch Engine

m – maskulinum

f - femininum

Nom.Sg. – nominativ singuláru

Part.Sg. – partitiv singuláru

Gen.Sg. – genitiv singuláru

Ines.Sg. – inessiv singuláru

Elat.Sg. – elativ singuláru

Ill.Sg. – illativ singuláru

Ades.Sg. – adesiv singuláru

Abl.Sg. – ablativ singuláru

All.Sg. – allativ singuláru

Tran.Sg. – translativ singuláru

Ess.Sg. – essiv singuláru

Nom.Pl. – nominativ plurálu

Part.Pl. – partitiv plurálu

Gen.Pl. – genitiv plurálu

Ines.Pl. – inessiv plurálu

Elat.Pl. – elativ plurálu

Ill.Pl. – illativ plurálu

Ades.Pl. – adessiv plurálu

Abl.Pl. – ablativ plurálu

All.Pl. – allativ plurálu

Tran.Pl. – translativ plurálu

Ess.Pl. – essiv plurálu

## Yhteenveto

Tämän diplomityön aiheena on ajan substantiivien kuvaus suomen ja tšekin kielessä. Päämenetelmänä on korpuslingvistiikka. Työn tavoitteena on ajan substantiivien tutkiminen käyttöyhteyksissään ja tutkimuksen tulosten perusteella hakusanojen luominen uutta suomi-tšekki-suomi sanakirjaa varten.

Työn teoreettisena lähtökohtana on kontekstuaalisen lingvistiikan näkökulma sanasemantiikkaan. Kontekstualismin isänä pidetään englantilaista kielitieteilijää J.R.R. Firthia, jonka mukaan tärkeänä piirteenä kaikilla kielen tasoilla on käyttöyhteys, eli konteksti (context of situation) (Firth, 1957). Tämä periaate on lähtökohtana myös korpuslingvistiikalle, joka isojen tekstikorpusten avulla mahdollistaa kielen tutkimusta laajoissa käyttöyhteyksissä – luonnollisissa teksteissä.

Kielikorpuksista on sanojen esiintymis- ja käyttöyhteysesimerkkien rinnalla mahdollista saada myös erilaisia tilastollisia tietoja, kuten mm. absoluuttinen ja relatiivinen frekvenssi, sanojen frekvenssi kielessä (useimmiten esiintyviä sanoja) ja tietoja sanojen kollokaatioista, sanojen tyypillisimmistä esiintymisistä konteksteissaan. Nykyiset kielikorpukset pystyvät laskentaohjelmien avulla suodattamaan teksteistä tyypillisiä sanaesiintymiä ja sanayhteyksiä, sanojen aktuaaliset merkitykset ovat sanojen semantiikan tärkeänä osana (konteksti-, kollokaatiomerkitys).

Tärkeää on myös kielikorpusten avulla saatujen tietojen käyttö. Kielikorpukset auttavat nykyään sanakirjatekijöitä, ne antavat tarkempia tietoja sanojen semantiikasta ja käytöstä ja tarjoavat myös luonnollisia käyttöesimerkkejä. Kielikorpukset voivat auttaa myös vieraiden kielten opiskelijoita antamalla nopeasti käyttöesimerkkejä kielen eri piirteistä.

Työn tutkimusmateriaalina on Tieteen tietotekniikan keskuksen (CSC) Kielipankki ja Tšekin kielen kansalliskorpuksen (ČNK, Český národní korpus, Czech National Corpus) eri korpuksset. Materiaaliaineistona on Kielipankin Helsingin Sanomien vuoden 1995 aineisto, joka sisältää yli 22 miljoonaa sanaa, ja ČNK:n SYN2000, SYN2005 ja SYN2006PUB korpuksset, joihin sisältyy yhteensä noin 500 miljoonaa sanaa. Kielipankin materiaaleja on mahdollista käyttää WWW-Lemmie 2.0 konkordanssiohjelman avulla, tämä on-line ohjelma osaa myös antaa tietoja frekvensseistä ja laskea kollokaatiotaulukoita, sama koskee myös ČNK:n korpuksia. Materiaalin käsittelyyn käytettiin kuitenkin myös Microsoftin Excel ja Access ohjelmia.

Ajan substantiivien ryhmä on sekä semanttisesti että frekvenssiltään tärkeä substantiivien ryhmä. Ajan substantiivit kuuluvat useimmiten esiintyvien sanojen joukkoon, Suomen kielen taajuussanaston (1979) mukaan substantiivi aika on 20., vuosi 23. ja päivä 51. useimmiten esiintyvä sana suomen kielessä. Ajan substantiivien asema kielessä on sen vuoksi vahva ja niillä on monia eri merkityksiä ja käyttötapauksia.

Korpuslingvistiikkamenetelmin tutkittiin substantiivien aika, vuosi, kuu, kuukausi, viikko, päivä, tunti, minuutti ja sekunti esiintymiä Kielipankin Helsingin Sanomien vuoden 1995 aineistossa ja vastaavien tšekin kielen substantiivien čas, rok, měsíc, týden, den, hodina, minuta, sekunda ja vteřina esiintymiä ČNK:n korpuksissa. Kontekstiesiintymisten rinnalla tutkittiin myös sanojen kollokaatioita ja semanttisia funktioita konteksteissaan. Michaela Mahlbergin (2005) mallin mukaan tarkasteltiin myös ajan substantiivien asemaa ”yleisten substantiivien” joukossa (”general nouns”). Mahlbergin mukaan ”yleisillä substantiiveilla” on sekä leksikaalisten että kieliopillisten yksiköiden ominaisuuksia, ja joissain tapauksissa ne voivat korvata deiktisiä sanoja.

Ajan substantiivien tarkastelussa tutkittiin ensinnäkin eri lekseemien frekvenssiä korpuksessa. Absoluuttisen frekvenssin lisäksi tutkittiin myös eri muotojen frekvenssiä. Sinclairin mukaan (1991) on tärkeä erottaa lekseemejä ja sanojen konkreettisia esiintymiä siinä mielessä, että lekseemit ovat vain abstraktisia yksiköitä, ja koska kontekstissa esiintyvät sanojen konkreettiset muodot (word), voimme sanamuotoja konteksteissaan pitää itsenäisinä yksikköinä.

Toiseksi tutkittiin eri muotojen esiintymiä korpuksessa ja verrattiin niiden syntaktisia ja semanttisia funktioita esiintymiskonteksteissaan, sekä niiden ekvivalentteja tšekin kielessä. Tämän luokittelun avulla ja eri tilastollisten tietojen ja kollokaatiotaulukkojen perusteella luotiin lopulta sanakirjahakusanoja, jotka voivat olla mallina uutta suomi-tšekki-suomi sanakirjaa varten. Nämä sanakirjahakusanat ovat vain ehdotuksia, sillä niiden lopullinen malli ja ulkoasu riippuisivat koko sanakirjan suunnitelmasta ja projektista.

Tämän työn tuloksena on todetus, että eri kielten kielikorpusten avulla on mahdollista saada riittäviä tietoja ja materiaaleja kielestä voidakseen kuvailla koko kielen tilannetta ja käyttää näitä tietoja kielestä sanakirjojen tekemiseen. Kielikorpukset antavat paitsi luotettavia ja luonnollisia tietoja kielestä myös sopivia sanojen käyttöesimerkkejä, jotka voivat auttaa sanakirjojen käyttäjiä kielen oppimisessa ja käytössä, sekä nopeuttavat ja täsmentävät sanakirjojen tekijöiden työtä. Muiden vanhempien menetelmien ja sanakirjojen rinnalla kielikorpukset ovat nykyään välttämättömiä keinoja kielen tulkimisessa sekä uusien kielikäsikirjojen ja sanakirjojen tekemisessä.

## **Literatura:**

**Altenberg, Bengt – Granger, Sylviane (Ed.): Lexis in Contrast. Corpus based approaches (2002).** Studies in Corpus Linguistics. John Benjamin Publishing Company. Amsterdam/Philadelphia.

**Atwell, Eric – Souter, Clive (Ed.): Corpus-based Computational Linguistics (1993).** Rodopi. Amsterdam/Atlanta.

**Bazell, C.E. – Catford, J.C. – Halliday, M.A.K. – Robins, R.H. (Ed.) In Memory of J.R. Firth (1966).** Longman. London.

**Blatná, Renáta – Čermák, František (Ed.): Manuál lexikografie (1995).** H&H. Praha.

**Čermák, František – Šulc, Michal (Ed.): Kolokace (2006).** Nakladatelství Lidové noviny, Praha.

**Encyklopedický slovník češtiny (2002).** Nakladatelství Lidové noviny. Praha.

**Firth, J. R.: Papers in Linguistics 1934-1951 (1957).** Oxford University Press. London.

**Grönroos, Mickel: The Lexical Database of The Bank of Finnish (2000).** Department of General Linguistics, University of Helsinki. Helsinki.

**Halliday, M.A.K. (1991): Corpus studies and probabilistic grammar.** In: Aijmer, Karin; Altenberg, Bengt (Ed.): English Corpus Linguistics. Studies in Honour of Jan Svartvik. Longman. London/New York.

**Halliday, M.A.K. – Hasan, Raquaiya: Cohesion in English (1976).** Longman. London.

**Church, K.W. – Hanks, P.: Word Association Norms, Mutual Information, and Lexicography (1990).** Computational Linguistics 16, s. 22-29.

**Iso suomen kielioppi (2004).** SKS. Helsinki.

**Kennedy, Graemme: An Introduction to Corpus Linguistics (1998).** Longman. London and New York.

**Kocek, Jan. – Kopřivová, Marie. – Kučera, Karel. (Ed.): Český národní korpus. Úvod a příručka uživatele (2000).** ÚČNK FF UK. Praha.

**Landau, Sidney I.: Dictionaries. The Art and Craft of Lexicography (2001).** Cambridge University Press. Cambridge.

**Lehtinen, Tapani: Suomen kielen tausta II. Suomi uralilaisena kielenä (2000).** Helsingin yliopisto. Helsinki.

**Lindroos, Hilikka Annikki: Finské postpozice (1997).** Univerzita Karlova. Praha.

**Lindroos, Hilikka: Suomi-tšekki-suomi taskusanakirja (1984).** WSOY. Helsinki.

**Mahlberg, Michaela: English General Nouns. A corpus theoretical approach (2005).** John Benjamin Publishing Company. Amsterdam/Philadelphia.

**McEnery, T. – Xiao, R. – Tono, Y. (2006): Corpus-Based Language Studies. An Advanced Resource Book.** Routledge. London/New York.

**Oakes, Michael P.: Statistics for Corpus Linguistics (1998).** Edinburgh University Press. Edinburgh.

**Ooi, Vincent B.Y.: Computer Corpus Lexicography (1998).** Edinburgh University Press. Edinburgh.

**Salminen, Anna: Minne menet aika? Suomen kielen konventionaalisten ajanilmausten fraseologista ja semanttiskognitiivista tarkastelua (2006).** Tampereen yliopisto. Tampere. (<http://tutkielmat.uta.fi/pdf/gradu01246.pdf>)

**Saukkonen, Pauli – Haipus, Marjatta – Niemikorpi, Antero – Sulkala, Helena: Suomen kielen taajuussanasto (1979).** WSOY. Porvoo, Helsinki, Juva.

**Short, Mick – Thomas, Jenny (Ed.): Using Corpora for Language Research (1996).** Longman. London/New York.

**Sinclair, John: Corpus, Concordance, Collocation (1991).** Oxford University Press. Oxford.

**Sinclair, John: The Lexical Item (1998).** In: Weigand, Edda (Ed.): Contrastive Lexical Semantics. John Benjamins Publishing Company. Amsterdam/Philadelphia.

**Slovník spisovné češtiny pro školu a veřejnost (2001).** Academia. Praha.

**Slovník spisovného jazyka českého (1989).** Academia. Praha.

**Sterkenburg, Piet van: A Practical Guide to Lexicography (2003).** John Benjamin Publishing Company. Amsterdam/Philadelphia.

**Suomen kielen perussanakirja (1990-1994).** Kotus. Helsinki.

**Šulc, Michal – Králík, Jan: The Representativeness of Czech Corpora (2005).** In: International Journal of Corpus Linguistics. Roč. 10, č. 3, s. 357-366.

**Šulc, Michal: Tematická reprezentativnost korpusů (2001).** In: Slovo a slovesnost. Roč. 62, č. 1, s. 53-61.

**Tognini-Bonelli, Elena: Corpus Linguistics at work (2001).** Studies in Corpus Linguistics. John Benjamin Publishing Company. Amsterdam/Philadelphia.



Český národní korpus - SYN2000. Ústav českého národního korpusu FF UK, Praha 2000. Dostupný z WWW: <http://ucnk.ff.cuni.cz> (listopad 2007 – duben 2008)

Český národní korpus - SYN2005. Ústav českého národního korpusu FF UK, Praha 2005. Dostupný z WWW: <http://ucnk.ff.cuni.cz> (listopad 2007 – duben 2008)

Český národní korpus - SYN2006PUB. Ústav českého národního korpusu FF UK, Praha 2006. Dostupný z WWW: <http://ucnk.ff.cuni.cz> (listopad 2007 – duben 2008)

Helsingin Sanomat 1995. An electronic document collection of the Finnish language containing 22 million words. Gatherers: The Department of General Linguistics, University of Helsinki; The University of Joensuu; CSC - Scientific Computing Ltd. Available through CSC, <http://www.csc.fi/>.